Information Hiding
An Application of Wiretap Channels with Side Information

Dissertation

zur Erlangung des Grades
eines Doktor-Ingenieurs

dem Fachbereich 12 - Maschinenwesen
der Universitaet Duisburg-Essen
vorgelegt von

Chaichana Mitrpant
aus Bangkok (Thailand)

Tag der muendlichen Pruefung: 27. November 2003

Vorsitzender der Pruefungskommission:	Prof. Dr.-Ing. Rudolf Tracht
Gutachter:	Prof. Dr.-Ing. A. J. Han Vinck
	Prof. Dr. Tran van Trung

Contents

Chapter 1

Introduction

Digital communications have evolved to become an essential part of everyday life since the groundbreaking work on information theory of Shannon in 1948 [17]. Coding theory is a major tool that enables one to realize the goal of having control over efficiency and reliability in a communication situation. High efficiency and high reliability are conflicting goals. To obtain high reliability, one often has to sacrifice efficiency. For instance, a channel coding strategy, consisting of codebook generating, encoding and decoding processes, generally introduces redundancy in a message to combat against noise in a channel so that a sender can transmit a message with high reliability. The introduced redundancy leads to low efficiency. A large amount of scientific work has been dedicated to analyzing and synthesizing different coding strategies so that the efficiency and the reliability can optimally be traded off to attain a set goal.

As digital communications become widely used, different constraints are imposed on the design of communication systems according to different user requirements. The demand to communicate sensitive messages calls for techniques that maintain the secrecy of the message in an efficient and reliable communication. Under certain circumstances, not only the message has to be kept secret, but also the communication itself has to be inconspicuous, which is known as covert communication.

1.1 Covert Communication

The covert communication scenario usually involves three parties: a sender, a recipient and an adversary. The sender is interested in covertly sending a message to the recipient with high efficiency and reliability through a channel while keeping the message unknown to the adversary. The goal of a system designer is to device a coding strategy that enables the optimal tradeoff between the three objectives.

A commonly used approach to maintain the secrecy of the message is to use key encryption, which transforms the message into a cryptogram unintelligible to the adversary, but decryptable with a key by the recipient to obtain the original message. If the adversary does not gain any

knowledge about the message from observing the cryptogram, the encryption algorithm then achieves the perfect secrecy. An example of the perfect secrecy encryption algorithm is the well-known one-time-pad algorithm. However, the perfect secrecy is impractical for the requirement that the length of the key must be at least the same as the length of the message, but it can serve as an upper bound on the level of secrecy quantified by equivocation. For practical reasons, computational security becomes a criterion for security of encryption algorithms.

An alternative approach to enhance the level of secrecy took by Wyner in [20] is to make an assumption that the adversary observes a degraded version of what the recipient observes through a wiretap channel. Assuming this wiretap channel model, a coding strategy without key can be used to encode the message by transforming it into a codeword so that the perfect secrecy can be achieved asymptotically.

The two approaches of attaining secrecy without any adaptation do not however provide the covert communication since the communication is not concealed by any mechanism albeit the transformation of the message. The second approach of using the wiretap channel model is taken in this thesis to be investigated and extended for the covert communication in an information theoretic framework.

1.2 Information Theoretic Framework and Challenges

A close examination of the wiretap channel model introduced by Wyner under the information theoretic framework reveals the fundamental ideas of coding strategies serving as a bedrock shared by the proofs of the achievabilities of channel capacities of the broadcast channel introduced by Cover in [4] and the dirty-paper channel introduced by Costa in [3]. The concept of code partitioning is used in the broadcast channel to transmit multiple messages to multiple recipients simultaneously in an efficient way. It is used in the dirty-paper channel to eliminate the effect of an interference known in advance as side information to the sender in a message transmission through a channel with interference. In the wiretap channel model, it is used to confuse the adversary to maintain the secrecy of the message from the sender intended for the recipient.

The code-partitioning concept exhibits a strong potential to be used in the covert communication due to its capability to confuse the adversary and mitigate the interference in the channel if it is known in advance to the sender. The challenge is to formulate a model suitable for the covert communication and apply the code-partitioning strategy to obtain a set of possible tradeoffs between efficiency, reliability and secrecy. A further challenge is to provide an implementation for the model.

To meet the challenges, we proposed an extension of the wiretap channel model to Gaussian wiretap channel with side information by taking advantage of the side information about the covert communication channel available to the sender in advance. We also investigate the role of side information when it is available to the adversary based on another variant of the wiretap channel, namely, the wiretap channel II.

1.3 Structure of the Thesis

We begin taking on the challenges by developing necessary information theoretic tools around the Asymptotic Equipartition Properties of sequences of independent identically distributed random variables in Chapter 2. The concept of code partitioning is described for the broadcast channel, the wiretap channel, the wiretap channel II, the Gaussian wiretap channel (another variant for the wiretap channel) and the dirty-paper channel in Chapter 3. The uses of the concept are compared and contrasted.

In Chapter 4, a new communication model is obtained by combining the Gaussian wiretap channel and the dirty-paper channel introduced in Chapter 3, and the code partitioning is employed to derive an achievable region. The new model is called Gaussian wiretap channel with side information for it consists of an additive white Gaussian noise main channel with additive white Gaussian interference and an additive white Gaussian noise wiretap channel. An appropriate level of interference can serve as a covert communication channel.

The implementation of the Gaussian wiretap channel with side information model as a system for information embedding in an image is proposed in Chapter 5. The image is used as a covert communication channel for transmission of a message intended for the recipient and to be kept secret from the adversary. Furthermore, it corresponds to the interference in the model, which is available to the sender in advance while the noise in the wiretap channel is used as a key given to the sender and the recipient beforehand. The system is implemented by enhancing the message secrecy of the watermarking scheme proposed by Eggers et. al in [8]. With an appropriate key, a high level of secrecy can be attained as shown in simulations.

The side information about the interference in the main channel in the Gaussian wiretap channel with side information is given to the sender in advance and can be used in the encoding process to mitigate the effect of the interference on the message transmission as in the case of the dirty-paper channel. However, if some side information is available to the adversary, the question of how it affects the secrecy of the message arises. In Chapter 6, we investigate this situation for the wiretap channel II described in Chapter 3. The model under this new circumstance is called wiretap channel II with side information since a part of the uncoded message is available to the adversary as side information, and it can be used to help his decoding for the message. The coset coding method with linear block codes used by Ozarow and Wyner in [16] to prove the capacity of the wiretap channel II is assumed in proving an achievable region for the wiretap channel II with side information. Moreover, some characteristics of good finite-length codes are derived for the model. Chapter 7 provides concluding remarks and possible directions for further research on the topic.

Chapter 2

Asymptotic Equipartition Properties

Basic definitions of information theoretic terms for discrete and continuous random variables and sequences are provided in this chapter for the development of tools crucial for understanding the code-partitioning concept and for the analysis of the Gaussian wiretap channel with side information (GWCSI) and the wiretap channel II with side information (WT2CSI). The tools are developed around the Asymptotic Equipartition Properties (AEP) of sequences of independent identically distributed (i.i.d.) random variables. The coding technique based on the AEP and a jointly typical decoder commonly used in the achievability proof of channel capacity is outlined for its frequent use in the thesis.

2.1 Basic Definitions

As usual, the measure of uncertainty of a random variable is entropy. The logarithm is of base 2 while the natural logarithm of x is denoted by $\ln(x)$. When we deal with Gaussian random variables, it is more convenient to use nat as a unit of information, where 1 nat is $\log(e)$ bits (about 1.44 bits). The definition of entropy of a discrete random variable and the definitions of joint entropy, conditional entropy and mutual information of a pair of discrete random variables are given below.

Definition 2.1 *The entropy $H(X)$ of a discrete random variable X with probability mass function $p(x)$ is defined as*

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x).$$

Definition 2.2 *The joint entropy $H(X,Y)$ of a pair of discrete random variables (X,Y) with joint probability mass function $p(x,y)$ is defined as*

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x,y).$$

Definition 2.3 *The conditional entropy $H(Y|X)$ of a pair of discrete random variables (X, Y) with joint probability mass function $p(x, y)$ and conditional probability mass function $p(y|x)$ is defined as*

$$H(Y|X) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y|x).$$

Definition 2.4 *The mutual information $I(X;Y)$ of a pair of discrete random variables (X, Y) with joint probability mass function $p(x, y)$ and marginal probability mass functions $p(x)$ and $p(y)$ is defined as*

$$I(X;Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}.$$

The corresponding definitions of differential entropy, joint differential entropy, conditional differential entropy and mutual information for continuous random variables are defined in the same ways by replacing the summations with integrations and (joint /conditional) probability mass functions with (joint /conditional) probability density functions. Note that we use the same notation for both entropy and differential entropy, and we must take special care when differential entropy is infinite and keep in mind that entropy is non-negative while differential entropy can be negative.

For a discrete memoryless channel whose input is X^N and output is Y^N with transition probability $p(y^N|x^N) = \prod_{i=1}^N p(y_i|x_i)$, the capacity of the channel is defined below.

Figure 2.1: The discrete memoryless channel with input X^N and output Y^N.

Definition 2.5 *The channel capacity of a discrete memoryless channel whose input is X and output is Y is defined as*

$$C = \sup_{p(x)} I(X;Y)$$

where the supremum is taken over all possible input probability mass functions $p(x)$.

From the above definitions, the relations among entropy, conditional entropy and mutual information listed below can be derived: (See [6].)

$$\begin{aligned}
H(X,Y) &= H(X) + H(Y|X) && (2.1)\\
&= H(Y) + H(X|Y) && (2.2)\\
H(X,Y) &= H(X,U,Y) - H(U|X,Y) && (2.3)\\
I(X;Y) &= H(X) - H(X|Y) && (2.4)\\
&= H(Y) - H(Y|X) && (2.5)\\
&= H(X) + H(Y) - H(X,Y) && (2.6)
\end{aligned}$$

Lemma 2.1 *Let (X^N, Y^N) be a pair of random vectors such that*

$$p(x^N, y^N) = \prod_{i=1}^{N} p(x_i, y_i).$$

Then $I(X^N; Y^N) = NI(X; Y)$.

Proof: If X^N and Y^N are discrete, by the definition of mutual information,

$$\begin{aligned} I(X^N; Y^N) &= \sum_{(x^N, y^N) \in \mathcal{X}^N \times \mathcal{Y}^N} p(x^N, y^N) \log \frac{p(x^N, y^N)}{p(x^N)p(y^N)} \\ &= \sum_{i=1}^{N} \sum_{(x^N, y^N) \in \mathcal{X}^N \times \mathcal{Y}^N} p(x^N, y^N) \log \frac{p(x_i, y_i)}{p(x_i)p(y_i)} \\ &= \sum_{i=1}^{N} \sum_{(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}} p(x_i, y_i) \log \frac{p(x_i, y_i)}{p(x_i)p(y_i)} \\ &= NI(X; Y). \end{aligned}$$

If X^N and Y^N are continuous, replace the summations with the integrations yielding the same result. ■

2.2 Gaussian Random Variables and Sequences

In this section, the definitions and the information theoretic properties are applied to Gaussian random variables (sequences) since they will be used in Chapter 4.

Differential Entropy.
For a zero-mean σ_X^2-variance Gaussian random variable $X \sim \mathcal{N}(0, \sigma_X^2)$ with probability density function

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{ -\frac{x^2}{2\sigma_X^2} \right\},$$

$$\begin{aligned} H(X) &= -\int_{\mathcal{X}} p(x) \ln p(x)\, dx \\ &= -\int_{\mathcal{X}} p(x) \left[-\frac{x^2}{2\sigma_X^2} - \frac{1}{2} \ln(2\pi\sigma_X^2) \right] dx \\ &= \frac{1}{2} + \frac{1}{2} \ln(2\pi\sigma_X^2) \\ &= \frac{1}{2} \ln(2\pi e\sigma_X^2) \text{ nats} \qquad (2.7) \\ &= \frac{1}{2} \log(2\pi e\sigma_X^2) \text{ bits.} \qquad (2.8) \end{aligned}$$

Note that $H(X) < 0$ if $2\pi e\sigma_X^2 < 1$. Furthermore, differential entropy of a Gaussian random variable depends on its variance, and it can theoretically be infinite. However, we only consider

power-limited Gaussian random processes; consequently, differential entropies considered here are bounded from above.

Joint Differential Entropy.
For a pair of Gaussian random variables $(X_1, X_2) \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$ with means zero and covariance matrix $\mathbf{K}$ and probability density function

$$p(\mathbf{x}) = \frac{1}{2\pi\sqrt{|\mathbf{K}|}} \exp\Big\{ -\frac{1}{2}\mathbf{x}^T\mathbf{K}^{-1}\mathbf{x}\Big\},$$

$$\begin{aligned}
H(X_1, X_2) &= -\int_{\mathcal{X}_1,\mathcal{X}_2} p(\mathbf{x}) \ln p(\mathbf{x})\ d\mathbf{x} \\
&= -\int_{\mathcal{X}_1,\mathcal{X}_2} p(\mathbf{x}) \Big[-\frac{1}{2}\mathbf{x}^T\mathbf{K}^{-1}\mathbf{x} - \frac{1}{2}\ln((2\pi)^2|\mathbf{K}|)\Big]\ d\mathbf{x} \\
&= -\int_{\mathcal{X}_1,\mathcal{X}_2} p(\mathbf{x}) \Big[-\frac{1}{2}\mathbf{x}^T\mathbf{K}^{-1}\mathbf{x}\Big]\ d\mathbf{x} + \frac{1}{2}\ln((2\pi)^2|\mathbf{K}|) \\
&= \frac{1}{2}E\Big[\sum_{i,j} X_i(\mathbf{K}^{-1})_{i,j}X_j\Big] + \frac{1}{2}\ln((2\pi)^2|\mathbf{K}|) \\
&= \frac{1}{2}\Big[\sum_j\sum_i \mathbf{K}_{j,i}(\mathbf{K}^{-1})_{i,j}\Big] + \frac{1}{2}\ln((2\pi)^2|\mathbf{K}|) \\
&= \frac{1}{2}\Big[\sum_j (\mathbf{K}\mathbf{K}^{-\mathbf{1}})_{j,j}\Big] + \frac{1}{2}\ln((2\pi)^2|\mathbf{K}|) \\
&= \frac{2}{2} + \frac{1}{2}\ln((2\pi)^2|\mathbf{K}|) \\
&= \frac{1}{2}\ln((2\pi e)^2|\mathbf{K}|) \text{ nats.}
\end{aligned}$$

Similarly, for a Gaussian random vector $(X_1, X_2, \ldots, X_N) \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$ with probability density function

$$p(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^N|\mathbf{K}|^{1/2}} \exp\Big\{ -\frac{1}{2}\mathbf{x}^T\mathbf{K}^{-1}\mathbf{x}\Big\},$$

$$H(X_1, \ldots, X_N) = \frac{1}{2}\ln((2\pi e)^N|\mathbf{K}|) \text{ nats.} \tag{2.9}$$

Conditional Differential Entropy.
The conditional entropy $H(Y|X)$ of a pair of continuous random variables (X, Y) with joint probability density function $p(x, y)$ and conditional probability density function $p(y|x)$ is defined as

$$H(Y|X) = -\int_{\mathcal{X},\mathcal{Y}} p(x, y) \log p(y|x)\ dy\ dx.$$

It can be expressed in terms of joint differential entropy and differential entropy, when all the terms are finite, $H(X|Y) = H(X, Y) - H(Y)$. In particular,

$$H(X|Y) = \frac{1}{2}\ln\left[\frac{(2\pi e)|\mathbf{K}_{XY}|}{\sigma_Y^2}\right] \text{ nats} \tag{2.10}$$

when X and Y are jointly Gaussian with covariance matrix $\mathbf{K}_{XY}$.

Mutual Information.
The mutual information $I(X;Y)$ of a pair of continuous random variables (X,Y) with joint probability density function $p(x,y)$ and marginal probability density functions $p(x)$ and $p(y)$ is defined as

$$I(X;Y) = \int_{\mathcal{X},\mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \, dy \, dx.$$

The mutual information between two Gaussian random variables X and Y – with a covariance matrix $\mathbf{K}_{XY}$, and $|\mathbf{K}_{XY}| = \sigma_X^2 \sigma_Y^2 (1-\varrho^2)$, where $\varrho = E[XY]/(\sigma_X \sigma_Y)$ is the correlation coefficient – is $I(X;Y) = H(X) + H(Y) - H(X,Y)$ when all the terms are finite; hence,

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) \\ &\overset{(1)}{=} \frac{1}{2} \ln \left[\frac{(2\pi e \sigma_X^2)(2\pi e \sigma_Y^2)}{(2\pi e)^2 |\mathbf{K}_{XY}|} \right] \\ &= \frac{1}{2} \ln \left[\frac{\sigma_X^2 \sigma_Y^2}{|\mathbf{K}_{XY}|} \right] \\ &= \frac{1}{2} \ln \left[\frac{1}{1-\varrho^2} \right] \text{ nats} \\ &= \frac{1}{2} \log \left[\frac{1}{1-\varrho^2} \right] \text{ bits.} \end{aligned} \tag{2.11}$$

(1) follows from equations (2.7) and (2.9).

In particular, for $Y = X + \eta$, where $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$ is independent of X, X and Y are jointly Gaussian with correlation coefficient

$$\begin{aligned} \varrho &= \frac{E[XY]}{\sigma_X \sigma_Y} \\ &= \frac{E[X(X+\eta)]}{\sigma_X \sigma_Y} \\ &= \frac{E[X^2]}{\sigma_X \sigma_Y} \\ &= \frac{\sigma_X}{\sigma_Y}, \end{aligned}$$

and the mutual information between X and Y becomes

$$\begin{aligned} I(X;Y) &= \frac{1}{2} \ln \left[\frac{1}{1-(\sigma_X^2/\sigma_Y^2)} \right] \\ &= \frac{1}{2} \ln \left[\frac{\sigma_Y^2}{\sigma_Y^2 - \sigma_X^2} \right] \\ &= \frac{1}{2} \ln \left[\frac{\sigma_X^2 + \sigma_\eta^2}{\sigma_\eta^2} \right] \text{ nats.} \qquad (2.12) \\ &= \frac{1}{2} \log \left[\frac{\sigma_X^2 + \sigma_\eta^2}{\sigma_\eta^2} \right] \text{ bits.} \qquad (2.13) \end{aligned}$$

However, if X and Y are independent, $|\mathbf{K}_{XY}| = \sigma_X^2\sigma_Y^2$, which implies that $I(X;Y) = 0$.

Additive White Gaussian Noise Channel.
Considering an Additive White Gaussian Noise (AWGN) channel model with the input-output relation: $Y^N = X^N + \eta^N$, where input X^N is drawn i.i.d. according to $\mathcal{N}(0, \sigma_X^2)$, and noise η^N, independent of the input, is drawn i.i.d. according to $\mathcal{N}(0, \sigma_\eta^2)$ as depicted in Figure 2.2, we can see that the channel is memoryless, and the output sequence is i.i.d. according to $\mathcal{N}(0, \sigma_X^2 + \sigma_\eta^2)$.

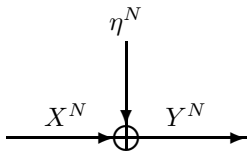

Figure 2.2: The Additive White Gaussian Noise Channel (AWGN) with noise $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$.

Since X^N, Y^N, η^N are i.i.d. and the channel is memoryless,

$$p(x^N, y^N) = \prod_{i=1}^{N} p(x_i, y_i),$$

and Lemma 2.1 can be applied, resulting in

$$\begin{aligned} I(X^N;Y^N) &= NI(X;Y) \\ &\overset{(1)}{=} \frac{N}{2}\ln\left[\frac{\sigma_X^2 + \sigma_\eta^2}{\sigma_\eta^2}\right] \text{ nats.} \end{aligned}$$

(1) follows from equation (2.12).

Now, for the AWGN depicted in Figure 2.2, if the average power of the input is limited to a constant P so that $\sigma_X^2 \leq P$, the channel capacity is defined as

$$C = \sup_{p(x):E[X^2]\leq P} I(X;Y) \tag{2.14}$$

where the supremum is taken over all possible input probability density functions $p(x)$ that satisfy the power constraint. Using the proof outlined in section 2.5, it is shown in [6] that the capacity of a P average-power-limited AWGN channel is achievable, where the capacity is

$$C_{AWGN} = \frac{1}{2}\log\left[\frac{P + \sigma_\eta^2}{\sigma_\eta^2}\right]. \tag{2.15}$$

In the next section, the Asymptotic Equipartition Properties, which are crucial for the proving technique outlined in section 2.5, are described.

2.3 Asymptotic Equipartition Properties

The Asymptotic Equipartition Properties (AEP) are concerned with the behaviors of sequences whose components are generated independently according to a probability distribution. It is a

direct consequence of the Weak Law of Large Numbers (WLLN); therefore, sequences or sets of sequences that are long enough exhibit these properties. (For the WLLN, see Theorem A.1 in Appendix A.) A sequence generated independently according to a probability distribution can be classified into the one that is typical and the one that is non-typical. The definition of typicality is given below. Here, we mainly use the AEP of continuous-alphabet sequences and will only consider the continuous case. For a set $T \subseteq \mathcal{R}^N$, the volume of the set T is denoted by $|T|$ and defined as

$$|T| = \int_T dx_1\ dx_2 \ldots dx_N.$$

Definition 2.6 *The set $T_X^N(\epsilon)$ of typical sequences x^N with respect to probability density function $p(x)$ is the set*

$$T_X^N(\epsilon) = \left\{ x^N \in \mathcal{X}^N : | -\frac{1}{N} \log p(x^N) - H(X)| < \epsilon \right\},$$

where $p(x^N) = \prod_{i=1}^N p(x_i)$.

A sequence x^N is said to be ϵ-typical with respect to function $p(x)$ if $x^N \in T_X^N(\epsilon)$. The probability that a typical sequence occurs is bounded in Lemma 2.2.

Lemma 2.2 *If $x^N \in T_X^N(\epsilon)$ with respect to probability density function $p(x)$, then*

$$2^{-N[H(X)+\epsilon]} < p(x^N) < 2^{-N[H(X)-\epsilon]}. \tag{2.16}$$

Proof: From the definition of typical set $T_X^N(\epsilon)$, we see that

$$\begin{array}{rcl} & | -\frac{1}{N} \log p(x^N) - H(X)| & < \epsilon \\ -\epsilon < & -\frac{1}{N} \log p(x^N) - H(X) & < \epsilon \\ H(X) - \epsilon < & -\frac{1}{N} \log p(x^N) & < H(X) + \epsilon \\ 2^{-N[H(X)+\epsilon]} < & p(x^N) & < 2^{-N[H(X)-\epsilon]} \end{array}$$

■

Lemma 2.2 implies that the typical sequences in $T_X^N(\epsilon)$ are roughly equally likely. Furthermore, if N, the length of the sequences, is large enough, probability that a sequence generated independently according to the probability density function $p(x)$ is in the typical set $T_X^N(\epsilon)$ is close to one as shown in Lemma 2.3.

Lemma 2.3 *For any $\epsilon > 0$ and $\epsilon_1 > 0$, there exists an integer N such that $Pr\{X^N \in T_X^N(\epsilon)\} > 1 - \epsilon_1$.*

Proof: Let $X' = -\log(p(X))$ with variance $\sigma^2_{X'}$. Applying the WLLN (Theorem A.1) to X',

$$\begin{aligned}
\Pr\left\{\left|\frac{1}{N}\sum_{i=1}^{N} X'_i - \int_{\mathcal{X}} -\log(p(x))p(x)\,dx\right| < \epsilon\right\} &> 1 - \frac{\sigma^2_{X'}}{N\epsilon^2} \\
\Pr\left\{\left|\frac{1}{N}\left[\sum_{i=1}^{N} -\log p(X_i)\right] - H(X)\right| < \epsilon\right\} &> 1 - \frac{\sigma^2_{X'}}{N\epsilon^2} \\
\Pr\left\{\left|\frac{1}{N}\left[-\log(\prod_{i=1}^{N} p(X_i))\right] - H(X)\right| < \epsilon\right\} &> 1 - \frac{\sigma^2_{X'}}{N\epsilon^2} \\
\Pr\left\{\left|-\frac{1}{N}\log(p(X^N)) - H(X)\right| < \epsilon\right\} &> 1 - \frac{\sigma^2_{X'}}{N\epsilon^2} \\
\Pr\left\{\left|-\frac{1}{N}\log(p(X^N)) - H(X)\right| < \epsilon\right\} &> 1 - \epsilon_1,
\end{aligned}$$

for $N > \sigma^2_{X'}/(\epsilon^2\epsilon_1)$. ■

The concept of typicality of a sequence can be extended to joint typicality of two sequences. Associated with the joint typicality, the joint AEP can be derived for a set of pairs of jointly typical sequences. These properties are the core of the idea used in the proof of the achievability of AWGN channel capacity outlined in section 2.5.

Definition 2.7 *The set $T^N_{X,Y}(\epsilon)$ of a jointly typical pair (x^N, y^N) with respect to joint probability density function $p(x, y)$ and marginal probability density functions $p(x)$ and $p(y)$ is the set*

$$\begin{aligned}
T^N_{X,Y}(\epsilon) = \Big\{&(x^N, y^N) \in \mathcal{X}^N \times \mathcal{Y}^N : \\
&|-\frac{1}{N}\log p(x^N) - H(X)| < \epsilon, && (2.17)\\
&|-\frac{1}{N}\log p(y^N) - H(Y)| < \epsilon, && (2.18)\\
&|-\frac{1}{N}\log p(x^N, y^N) - H(X,Y)| < \epsilon\Big\}, && (2.19)
\end{aligned}$$

where

$$p(x^N, y^N) = \prod_{i=1}^{N} p(x_i, y_i).$$

Based on the definition of the jointly typical set, various properties of members of the set and the set can be obtained. Lemma 2.4 bounds the probability, joint probability and conditional probability of the members of the set. Lemma 2.5 indicates that the probability that a pair of sequences (x^N, y^N) is in the jointly typical set $T^N_{X,Y}(\epsilon)$ is close to one for sufficiently large N. Finally, the volume of a jointly typical set is estimated in Lemma 2.6.

Lemma 2.4 *If $(x^N, y^N) \in T^N_{X,Y}(\epsilon)$ with respect to probability density function $p(x, y)$ and marginal probability density functions $p(x)$ and $p(y)$, then*

$$\begin{aligned}
2^{-N[H(X)+\epsilon]} &< p(x^N) &< 2^{-N[H(X)-\epsilon]},\\
2^{-N[H(Y)+\epsilon]} &< p(y^N) &< 2^{-N[H(Y)-\epsilon]},\\
2^{-N[H(X,Y)+\epsilon]} &< p(x^N, y^N) &< 2^{-N[H(X,Y)-\epsilon]},\\
2^{-N[H(Y|X)+2\epsilon]} &< p(y^N|x^N) &< 2^{-N[H(Y|X)-2\epsilon]}.
\end{aligned}$$

Proof: Use the definition of the joint typicality as in the proof of Lemma 2.2. ■

Lemma 2.5 *For any $\epsilon > 0$ and $\epsilon_1 > 0$, there exists an integer N such that $Pr\{(X^N, Y^N) \in T^N_{X,Y}(\epsilon)\} > 1 - \epsilon_1$.*

Proof: Apply the proof of Lemma 2.3 with $X' = -\log(p(X))$, $X' = -\log(p(Y))$, and $X' = -\log(p(X, Y))$ to the conditions (2.17), (2.18), and (2.19), respectively, and select N large enough so that all three conditions are satisfied. ■

Lemma 2.6 *For any $\epsilon > 0$, there exists an integer N such that $(1 - \epsilon)2^{N[H(X,Y)-\epsilon]} < |T^N_{X,Y}(\epsilon)| < 2^{N[H(X,Y)+\epsilon]}$*

Proof:

$$\begin{aligned}
1 &= \int_{\mathcal{X}^N \times \mathcal{Y}^N} p(x^N, y^N)\, dx^N\, dy^N\\
&> \int_{T^N_{X,Y}(\epsilon)} p(x^N, y^N)\, dx^N\, dy^N\\
&\overset{(1)}{>} \int_{T^N_{X,Y}(\epsilon)} 2^{-N[H(X,Y)+\epsilon]}\, dx^N\, dy^N\\
&= |T^N_{X,Y}(\epsilon)| 2^{-N[H(X,Y)+\epsilon]}\\
|T^N_{X,Y}(\epsilon)| &< 2^{N[H(X,Y)+\epsilon]},
\end{aligned}$$

and

$$\begin{aligned}
1 - \epsilon &\overset{(2)}{<} \Pr\{(X^N, Y^N) \in T^N_{X,Y}(\epsilon)\}\\
&= \int_{T^N_{X,Y}(\epsilon)} p(x^N, y^N)\, dx^N\, dy^N\\
&\overset{(3)}{<} \int_{T^N_{X,Y}(\epsilon)} 2^{-N[H(X,Y)-\epsilon]}\, dx^N\, dy^N\\
&= |T^N_{X,Y}(\epsilon)| 2^{-N[H(X,Y)-\epsilon]}\\
|T^N_{X,Y}(\epsilon)| &> (1 - \epsilon)2^{N[H(X,Y)-\epsilon]}.
\end{aligned}$$

(1) follows from Lemma 2.4. (2) follows from Lemma 2.5 for $\epsilon_1 = \epsilon$ and sufficiently large N. (3) follows from Lemma 2.4. ■

Lemma 2.5 indicates the probability of a pair of sequences generated i.i.d. according to the joint probability density function $p(x, y)$ being in $T^N_{X,Y}(\epsilon)$, which is used to calculate a part of probability of error in the proof outlined in section 2.5. The other part of the probability of error is calculated based on Lemma 2.7 stated below.

Lemma 2.7 *Let $(\tilde{X}^N, \tilde{Y}^N)$ be a pair of length-N sequences such that*

$$p(\tilde{x}^N, \tilde{y}^N) = \prod_{i=1}^{N} p(\tilde{x}_i, \tilde{y}_i),$$

with marginal probability density functions $p(\tilde{x})$ and $p(\tilde{y})$. If X^N and Y^N are two independent i.i.d. sequences of random variables with the marginal probability density functions $p(\tilde{x})$ and $p(\tilde{y})$, respectively, then

$$Pr\{(X^N, Y^N) \in T^N_{\tilde{X},\tilde{Y}}(\epsilon)\} < 2^{-N[I(\tilde{X};\tilde{Y})-3\epsilon]} \tag{2.20}$$

$$Pr\{(X^N, Y^N) \in T^N_{\tilde{X},\tilde{Y}}(\epsilon)\} > (1-\epsilon)2^{-N[I(\tilde{X};\tilde{Y})+3\epsilon]}, \tag{2.21}$$

for sufficiently large N.

Proof:

$$\begin{aligned}
\Pr\{(X^N, Y^N) \in T^N_{\tilde{X},\tilde{Y}}(\epsilon)\} &= \int_{T^N_{\tilde{X},\tilde{Y}}(\epsilon)} p(x^N, y^N)\, dx^N\, dy^N \\
&\overset{(1)}{=} \int_{T^N_{\tilde{X},\tilde{Y}}(\epsilon)} p(x^N)p(y^N)\, dx^N\, dy^N \\
&= \int_{T^N_{\tilde{X},\tilde{Y}}(\epsilon)} p(\tilde{x}^N)p(\tilde{y}^N)\, dx^N\, dy^N \\
&\overset{(2)}{>} \int_{T^N_{\tilde{X},\tilde{Y}}(\epsilon)} 2^{-N[H(\tilde{X})+\epsilon]}2^{-N[H(\tilde{Y})+\epsilon]}\, dx^N\, dy^N \\
&\overset{(3)}{>} (1-\epsilon)2^{N[H(\tilde{X},\tilde{Y})-\epsilon]}2^{-N[H(\tilde{X})+\epsilon]}2^{-N[H(\tilde{Y})+\epsilon]} \\
&= (1-\epsilon)2^{-N[-H(\tilde{X},\tilde{Y})+H(\tilde{X})+H(\tilde{Y})+3\epsilon]} \\
&\overset{(4)}{=} (1-\epsilon)2^{-N[I(\tilde{X};\tilde{Y})+3\epsilon]}
\end{aligned}$$

and

$$\begin{aligned}
\Pr\{(X^N, Y^N) \in T^N_{\tilde{X},\tilde{Y}}(\epsilon)\} &\overset{(5)}{<} \int_{T^N_{\tilde{X},\tilde{Y}}(\epsilon)} 2^{-N[H(\tilde{X})-\epsilon]}2^{-N[H(\tilde{Y})-\epsilon]}\, dx^N\, dy^N \\
&\overset{(6)}{<} 2^{N[H(\tilde{X},\tilde{Y})+\epsilon]}2^{-N[H(\tilde{X})-\epsilon]}2^{-N[H(\tilde{Y})-\epsilon]} \\
&= 2^{-N[-H(\tilde{X},\tilde{Y})+H(\tilde{X})+H(\tilde{Y})-3\epsilon]} \\
&\overset{(7)}{=} 2^{-N[I(\tilde{X};\tilde{Y})-3\epsilon]}
\end{aligned}$$

(1) follows from the independence of X^N and Y^N. (2) and (5) follow from Lemma 2.4. (3) and (6) follow from Lemma 2.6 for sufficiently large N. (4) and (7) follow from equation (2.6). ■

Lemma 2.7 gives an estimate of the probability of two sequences generated independently of each other being jointly typical when each of the sequences is generated i.i.d. according to the corresponding marginal probability density function associated with the jointly typical set $T^N_{\tilde{X},\tilde{Y}}(\epsilon)$.

2.4 AEP of Gaussian Sequences

In this section, we consider the AEP of Gaussian sequences and draw some characteristics of a typical Gaussian sequence, a set of pairs of jointly typical Gaussian sequences and its members. Lemma 2.8 indicates that a typical Gaussian sequence has average power approximately equal to the variance of the governing distribution. Consequently, a constraint on the average power of a typical Gaussian sequence can be imposed by specifying the governing distribution as shown in Lemma 2.9

Lemma 2.8 *If X^N is a sequence of random variables i.i.d. according to $\mathcal{N}(0,\sigma_X^2)$, and $x^N \in T_X^N(\epsilon)$ for any $\epsilon > 0$, then*

$$\left| \frac{\langle x^N, x^N \rangle}{N\sigma_X^2} - 1 \right| < 2\epsilon \ln(2).$$

Proof: From Definition 2.6,

$$\begin{aligned}
\epsilon &> \left| -\frac{1}{N}\log p(x^N) - H(X) \right| \\
&\overset{(1)}{=} \left| -\frac{1}{N}\sum_{i=1}^{N}\log p(x_i) - \frac{1}{2}\log(2\pi e\sigma_x^2) \right| \\
&\overset{(2)}{=} \frac{1}{\ln(2)}\left| \frac{1}{N}\sum_{i=1}^{N}\left[\frac{x_i^2}{2\sigma_x^2} + \frac{1}{2}\ln(2\pi\sigma_x^2)\right] - \frac{1}{2}\ln(2\pi e\sigma_x^2) \right| \\
&= \frac{1}{\ln(2)}\left| \frac{1}{N}\sum_{i=1}^{N}\frac{x_i^2}{2\sigma_x^2} - \frac{1}{2} \right| \\
&= \frac{1}{\ln(2)}\left| \frac{\langle x^N, x^N \rangle}{2N\sigma_x^2} - \frac{1}{2} \right| \\
2\epsilon\ln(2) &> \left| \frac{\langle x^N, x^N \rangle}{N\sigma_x^2} - 1 \right|
\end{aligned}$$

(1) follows from equation (2.8). (2) follows from the definition of the Gaussian probability density function. ■

Lemma 2.9 *Let X^N be a sequence of i.i.d. random variables $X \sim \mathcal{N}(0,\sigma_X^2)$. If $x^N \in T_X^N(\epsilon)$, for any $\epsilon > 0$, and $\sigma_X^2 \leq \frac{P}{1+2\epsilon\ln(2)}$, then $\frac{1}{N}\sum_{i=1}^{N} x_i^2 \leq P$.*

Proof: Since $x^N \in T_X^N(\epsilon)$, applying Lemma 2.8 to x^N yields

$$\begin{aligned}
\left| \frac{\langle x^N, x^N \rangle}{N\sigma_X^2} - 1 \right| &< 2\epsilon \ln(2) \\
\frac{\langle x^N, x^N \rangle}{N} &< \sigma_X^2 [1 + 2\epsilon \ln(2)] \\
&\overset{(1)}{\leq} \frac{P}{1 + 2\epsilon \ln(2)} [1 + 2\epsilon \ln(2)] \\
\frac{1}{N} \sum_{i=1}^{N} x_i^2 &\leq P.
\end{aligned}$$

(1) follows from the assumption on σ_X^2. ■

The result of Lemma 2.8 is known as the sphere hardening effect, which has a geometrical interpretation that normalized vector $\underline{x}^N = x^N/\sqrt{N}$ lies in the shell bounded by the solid lines depicted in Figure 2.3 if x^N is a typical Gaussian sequence.

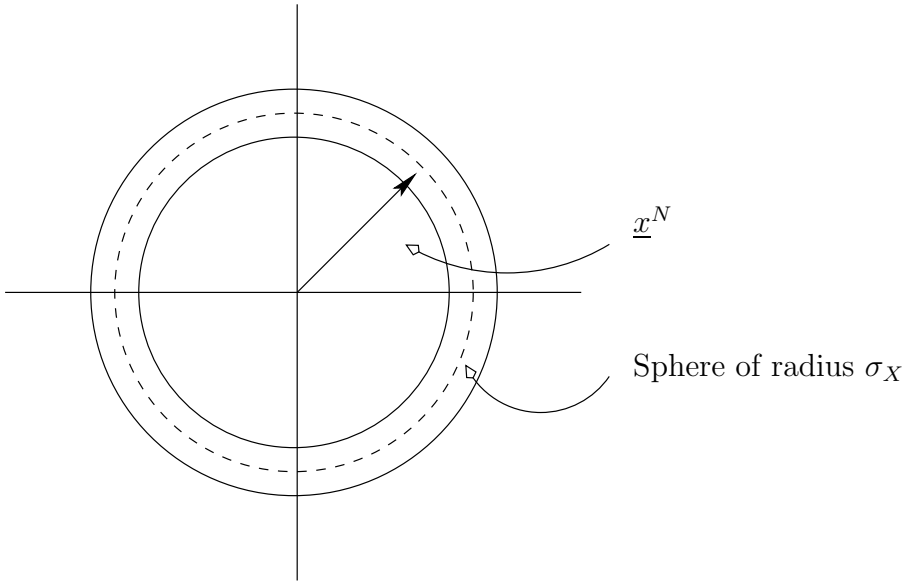

Figure 2.3: The locations of normalized typical Gaussian sequences.

An analogue of Lemma 2.8 for a pair of jointly typical Gaussian sequence is given in Lemma 2.10 below.

Lemma 2.10 *Let (X^N, Y^N) be a pair of sequences of random variables i.i.d. according to joint probability density function $p(x, y)$, and marginal probability density functions $p(x) \sim \mathcal{N}(0, \sigma_X^2)$, and $p(y) \sim \mathcal{N}(0, \sigma_Y^2)$, i.e.,*

$$p(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\varrho^2}} \exp\left\{ -\frac{1}{2(1-\varrho^2)} \left[\frac{x^2}{\sigma_X^2} - \frac{2\varrho xy}{\sigma_X\sigma_Y} + \frac{y^2}{\sigma_Y^2} \right] \right\},$$

where $\varrho = E[XY]/(\sigma_X\sigma_Y)$ *is an associated correlation coefficient. If* $(x^N, y^N) \in T^N_{X,Y}(\epsilon)$ *for any* $\epsilon > 0$, *then*

$$\begin{aligned}
\left|\frac{\langle x^N, x^N\rangle}{N\sigma_X^2} - 1\right| &< 2\epsilon\ln(2),\\
\left|\frac{\langle y^N, y^N\rangle}{N\sigma_Y^2} - 1\right| &< 2\epsilon\ln(2),\\
\left|\frac{\langle x^N, y^N\rangle}{N\varrho\sigma_X\sigma_Y} - 1\right| &< (\frac{3}{\varrho^2} - 1)\epsilon\ln(2).
\end{aligned}$$

Proof: The first two inequalities follows from applying Lemma 2.8 to the conditions (2.17) and (2.18), respectively. Considering the condition (2.19) with $|\mathbf{K}_{XY}| = \sigma_X^2\sigma_Y^2(1-\varrho^2)$, we have

$$\begin{aligned}
\epsilon &> \left| -\frac{1}{N}\log p(x^N, y^N) - H(X,Y)\right|\\
&\overset{(1)}{=} \left| -\frac{1}{N}\sum_{i=1}^N \log p(x_i, y_i) - \frac{1}{2}\log((2\pi e)^2|\mathbf{K}_{XY}|)\right|\\
&= \frac{1}{\ln(2)}\left|\frac{1}{N}\sum_{i=1}^N\left[\frac{1}{2(1-\varrho^2)}\left(\frac{x_i^2}{\sigma_X^2} - \frac{2\varrho x_i y_i}{\sigma_X\sigma_Y} + \frac{y_i^2}{\sigma_Y^2}\right)\right.\right.\\
&\quad \left.\left. + \ln(2\pi\sigma_X\sigma_Y\sqrt{1-\varrho^2})\right] - \ln(2\pi e\sigma_X\sigma_Y\sqrt{1-\varrho^2})\right|\\
&= \frac{1}{\ln(2)}\left|\frac{1}{N}\sum_{i=1}^N\left[\frac{1}{2(1-\varrho^2)}\left(\frac{x_i^2}{\sigma_X^2} - \frac{2\varrho x_i y_i}{\sigma_X\sigma_Y} + \frac{y_i^2}{\sigma_Y^2}\right)\right]\right.\\
&\quad \left. + \ln(2\pi\sigma_X\sigma_Y\sqrt{1-\varrho^2}) - \ln(2\pi e\sigma_X\sigma_Y\sqrt{1-\varrho^2})\right|\\
&= \frac{1}{\ln(2)}\left|\frac{1}{N}\sum_{i=1}^N\left[\frac{1}{2(1-\varrho^2)}\left(\frac{x_i^2}{\sigma_X^2} - \frac{2\varrho x_i y_i}{\sigma_X\sigma_Y} + \frac{y_i^2}{\sigma_Y^2}\right)\right] - 1\right|\\
&= \frac{1}{(1-\varrho^2)\ln(2)}\left|\frac{1}{2N}\sum_{i=1}^N\frac{x_i^2}{\sigma_X^2} + \frac{1}{2N}\sum_{i=1}^N\frac{y_i^2}{\sigma_Y^2} - \frac{1}{N}\sum_{i=1}^N\frac{\varrho x_i y_i}{\sigma_X\sigma_Y} - (1-\varrho^2)\right|\\
&= \frac{1}{(1-\varrho^2)\ln(2)}\left|\left[\frac{\langle x^N, x^N\rangle}{2N\sigma_X^2} - \frac{1}{2}\right] + \left[\frac{\langle y^N, y^N\rangle}{2N\sigma_Y^2} - \frac{1}{2}\right] - \left[\frac{\varrho\langle x^N, y^N\rangle}{N\sigma_X\sigma_Y} - \varrho^2\right]\right|.
\end{aligned}$$

(1) follows from equation (2.9) for two random variables. Applying Lemma 2.8 to the last inequality above, we have

$$\begin{aligned}
\left|\frac{\varrho\langle x^N, y^N\rangle}{N\sigma_X\sigma_Y} - \varrho^2\right| &< (1-\varrho^2)\epsilon\ln(2) + 2\epsilon\ln(2)\\
\left|\frac{\langle x^N, y^N\rangle}{N\varrho\sigma_X\sigma_Y} - 1\right| &< (\frac{3}{\varrho^2} - 1)\epsilon\ln(2).
\end{aligned}$$

■

In Lemma 2.11 and Lemma 2.12, a special setup, in which $U^N = X^N + \alpha V^N$ for a real constant α, and X^N and V^N are independent Gaussian sequences, is considered. This setup will be useful in the coding strategy used to mitigate an interference in the channel as employed in Chapter 4.

Lemma 2.11 *Let X^N and V^N be two independent sequences of i.i.d. random variables $X \sim \mathcal{N}(0, \sigma_X^2)$, and $V \sim \mathcal{N}(0, \sigma_V^2)$, respectively. Let $U^N = X^N + \alpha V^N$ for a constant real number α. If $(u^N, v^N) \in T_{U,V}^N(\epsilon)$, for any $\epsilon > 0$, then $u^N - \alpha v^N = x^N \in T_X^N(2\epsilon)$, and*

$$\left| \frac{\langle x^N, v^N \rangle}{N} \right| < \frac{\epsilon \ln(2)}{\alpha} [3\sigma_X^2 + 2\alpha^2 \sigma_V^2].$$

Proof: Since X^N and V^N are two independent sequences of i.i.d. Gaussian random variables, U^N is a sequence of i.i.d. Gaussian random variables drawn according to $\mathcal{N}(0, \sigma_X^2 + \alpha^2 \sigma_V^2)$. (See [7] and [10].)

Applying the condition (2.19) under the assumption that $(u^N, v^N) \in T_{U,V}^N(\epsilon)$ to obtain

$$\begin{aligned}
\epsilon &\overset{(1)}{>} \left| -\frac{1}{N} \log p(u^N, v^N) - H(U,V) \right| \\
&\overset{(2)}{=} \left| -\frac{1}{N} \log p(u^N, v^N) - H(V) - H(U|V) \right| \\
&\overset{(3)}{=} \left| -\frac{1}{N} \log p(u^N, v^N) - H(V) - H(X) \right| \\
2\epsilon &\overset{(1)}{>} \left| -\frac{1}{N} \log p(u^N, v^N) + \frac{1}{N} \log p(v^N) - H(X) \right| \\
&= \left| -\frac{1}{N} \log p(u^N | v^N) - H(X) \right| \\
&= \left| -\frac{1}{N} \log p(x^N) - H(X) \right|.
\end{aligned}$$

(1) follows from Definition 2.7. (2) follows from equation (2.2). (3) follows from the independence between X and V.

To show the second part of the Lemma, we begin with the fact that $x^N \in T_X^N(2\epsilon)$. By Lemma 2.8,

$$\begin{aligned}
4\epsilon \ln(2) &> \left| \frac{\langle x^N, x^N \rangle}{N\sigma_X^2} - 1 \right| \\
4\sigma_X^2 \epsilon \ln(2) &> \left| \frac{\langle u^N - \alpha v^N, u^N - \alpha v^N \rangle - N\sigma_X^2}{N} \right| \\
&= \left| \frac{\langle u^N, u^N \rangle + \alpha^2 \langle v^N, v^N \rangle - 2\alpha \langle u^N, v^N \rangle - N\sigma_X^2}{N} \right| \\
&= \left| \frac{\langle u^N, u^N \rangle + \alpha^2 \langle v^N, v^N \rangle - 2\alpha \langle x^N, v^N \rangle - 2\alpha^2 \langle v^N, v^N \rangle - N\sigma_X^2}{N} \right| \\
&= \left| \left[\frac{\langle u^N, u^N \rangle - N\sigma_U^2}{N} \right] - \left[\frac{\alpha^2 \langle v^N, v^N \rangle - N\alpha^2\sigma_V^2}{N} \right] \right. \\
&\quad \left. - \left[\frac{2\alpha \langle x^N, v^N \rangle - N\sigma_U^2 + N\alpha^2\sigma_V^2 + N\sigma_X^2}{N} \right] \right|.
\end{aligned}$$

Applying Lemma 2.8 to the first two terms in the sum to obtain

$$\begin{aligned}
4\sigma_X^2 \epsilon \ln(2) + 2\sigma_U^2 \epsilon \ln(2) + 2\alpha^2 \sigma_V^2 \epsilon \ln(2) &> \left| \left[\frac{2\alpha \langle x^N, v^N \rangle - N\sigma_U^2 + N\alpha^2\sigma_V^2 + N\sigma_X^2}{N} \right] \right| \\
2\epsilon \ln(2)[2\sigma_X^2 + \sigma_U^2 + \alpha^2\sigma_V^2] &\overset{(4)}{>} \left| \left[\frac{2\alpha \langle x^N, v^N \rangle}{N} \right] \right| \\
\frac{\epsilon \ln(2)}{\alpha}[3\sigma_X^2 + 2\alpha^2\sigma_V^2] &> \left| \frac{\langle x^N, v^N \rangle}{N} \right|.
\end{aligned}$$

(4) follows from the fact that $\sigma_U^2 = \sigma_X^2 + \alpha^2\sigma_V^2$. ■

Note that the second part of Lemma 2.11 implies that the normalized versions of the vectors x^N and v^N are almost orthogonal.

Lemma 2.12 *Let X^N and V^N be two independent sequences of i.i.d. random variables $X \sim \mathcal{N}(0, \sigma_X^2)$, and $V \sim \mathcal{N}(0, \sigma_V^2)$, respectively. Let $U^N = X^N + \alpha V^N$ for a constant real number α. If $(u^N, v^N) \in T_{U,V}^N(\epsilon)$, for any $\epsilon > 0$, and $\sigma_X^2 \leq \frac{P}{1+4\epsilon \ln(2)}$, then $\frac{1}{N}\sum_{i=1}^N x_i^2 \leq P$.*

Proof: Applying Lemma 2.11 to the pair (u^N, v^N) implies that $x^N \in T_X^N(2\epsilon)$. Applying Lemma 2.8 to x^N with 2ϵ yields

$$\begin{aligned}
\left| \frac{\langle x^N, x^N \rangle}{N\sigma_X^2} - 1 \right| &< 4\epsilon \ln(2) \\
\frac{\langle x^N, x^N \rangle}{N} &< \sigma_X^2 [1 + 4\epsilon \ln(2)] \\
&\leq \frac{P}{1 + 4\epsilon \ln(2)} [1 + 4\epsilon \ln(2)] \\
\frac{1}{N} \sum_{i=1}^{N} x_i^2 &\leq P.
\end{aligned}$$

■

2.5 Achievability of the AWGN Channel Capacity

The proof of the achievability of the average power constraint AWGN channel given in [6] is outlined here. It uses the random coding strategy with joint typicality decoding. The channel is depicted in Figure 2.2 with the average power constraint on the input:

$$\frac{1}{N} \sum_{i=1}^{N} x_i^2 \leq P.$$

The coding strategy is described below:

1. *Generating the codebook.* Generate a codebook consisting of 2^{NR} codewords X_c with each element i.i.d. according to a normal distribution with variance $P/(1 + 2\epsilon \ln(2))$, where R is the rate of transmission. Uniquely associate a message w with the codeword $X_c(w)$. Distribute the codebook to the sender and the recipient.

2. *Encoding.* To send a message w, transmit the codeword $X_c(w)$ in the codebook.

3. *Decoding.* The recipient looks for a codeword in the codebook that is jointly typical with the received vector. If there is one and only one such codeword, declare it to be the transmitted codeword. Otherwise, declare an error. An error is also declared if the codeword does not satisfy the power constraint.

A jointly typical decoder is not optimal in term of minimizing the probability of error, but it is simple to analyze and achieves all rates below capacity. The decoder works based on the AEP of long sequences of i.i.d. random variables. We note that there are two types of error that can occur in the above process. The first type of error occurs when the codeword does not satisfy the power constraint, and the second type of error occurs during the decoding process. The decoding error may happen when the decoder cannot identify the transmitted codeword from the received vector or when a wrong codeword is identified as the transmitted codeword.

In the achievability proof of the channel capacity, the probability distribution governing the input to the channel is chosen to be the channel capacity achieving distribution, which is the Gaussian distribution in case of AWGN channel. In this way, the channel capacity is achieved. To cope with the error events in the coding process, the channel is utilized repeatedly N times. When N is allowed to be sufficiently large, WLLN comes into effect, and the sequences considered in the coding scheme are very likely to be typical sequences. Lemma 2.3 and Lemma 2.9 are used to bound the probability of error of the first type. Lemma 2.5 and Lemma 2.7 are used to bound the probability of error of the second type. Given a required probability of error, the length of the sequences N is bounded from below via the four lemmas.

Chapter 3

Code Partitioning and Its Applications

3.1 Introduction

In a basic scenario of digital communication, a channel code is designed to achieve a reliable communication between a sender and a recipient at a high rate. The code described in section 2.5 of the previous chapter is, for instance, designed for a reliable communication via the AWGN channel with average input power constraint between a sender and a recipient at the rate close to the channel capacity. As the communication scenario becomes more sophisticated, the coding strategy must be adapted accordingly.

In this chapter, we investigate the concept and the roles of code partitioning in three situations: the broadcast channels in section 3.2, the wiretap channels in sections 3.3, 3.4 and 3.5, and the dirty-paper channel in section 3.6. The uses of the code-partitioning concept in the three situations are subsequently compared and contrasted in section 3.7. This concept will be used in the coding strategy for the Gaussian wiretap channel with side information to be introduced in Chapter 4.

3.2 Broadcast Channels

The broadcast channels, introduced in [4] by Cover, involves one sender and multiple recipients. The sender is interested in simultaneously and reliably sending multiple messages to multiple recipients at high rates via a broadcast channel, which takes one input and produces multiple outputs, each of which corresponds to a recipient. For simplicity, the two-recipient discrete-time memoryless broadcast channel defined in [5] is considered and simply referred to as broadcast channel.

The broadcast channel consists of an input alphabet $\mathcal{X}$ and two output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$

and probability transition function $p(y_1^N, y_2^N|x^N)$ such that

$$p(y_1^N, y_2^N|x^N) = \prod_{i=1}^{N} p(y_{1i}, y_{2i}|x_i).$$

Assuming that the decoding at the two recipients are done independently, only the knowledge of the marginal distributions, $p_1(y_1|x)$ and $p_2(y_2|x)$, are required.

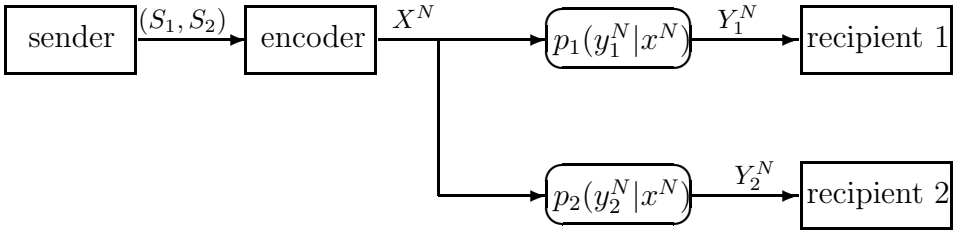

Figure 3.1: The broadcast channel.

Let (S_1, S_2) be a pair of independent messages, with S_1 intended for recipient 1 and S_2 intended for recipient 2, selected uniformly over $\{1, 2, \ldots, 2^{NR_1}\} \times \{1, 2, \ldots, 2^{NR_2}\}$. The message pair is encoded into a codeword X^N as the input of the broadcast channel using an $(N, (2^{NR_1}, 2^{NR_2}))$ code for the broadcast channel with independent messages. Hence, R_1 and R_2 are the rates of transmission to recipients 1 and 2, respectively. Recipient 1 decodes Y_1^N as $\hat{S}_1$, and recipient 2 decodes Y_2^N as $\hat{S}_2$. An error occurs if at least one of the recipients decodes incorrectly, i.e.,

$$P_e = \sum_{(s_1,s_2)\in \mathcal{S}_1 \times \mathcal{S}_2} \Pr\{(S_1, S_2) = (s_1, s_2)\} \Pr\{\hat{S}_1 \neq s_1 \text{ or } \hat{S}_2 \neq s_2\}.$$

A rate pair (R_1, R_2) is said to be achievable for the broadcast channel if there exists a sequence of $(N, (2^{NR_1}, 2^{NR_2}))$ codes with $P_e \to 0$ as $N \to \infty$, and the capacity region of the broadcast channel is the closure of the set of all achievable rates.

The coding strategy known as time-sharing strategy makes use of two codes: C^1 designed for channel $p_1(y_1|x)$ and C^2 designed for channel $p_2(y_2|x)$, and time slot allocation to the two recipients. In the time slots allocated to recipient 1, the code C^1 is used to encode S_1, and in the time slots allocated to recipient 2, the code C^2 is used to encode S_2. By varying the time slot allocation and the codes, an achievable region for the this strategy can be derived.

An alternative coding strategy for superimposing messages via code partitioning was proposed by Cover in [4] to enlarge the achievable region associated with the time-sharing strategy. The use of the code partitioning for broadcast channel is most salient in the degraded case of the broadcast channel, whose transition probability can be expressed as

$$p(y_2|x) = \sum_{y_1 \in \mathcal{Y}_1} p(y_1|x)\tilde{p}(y_2|y_1). \tag{3.1}$$

Two elucidating examples of degraded broadcast channels are given here to illustrate the use of the code partitioning to superimpose messages.

Binary Symmetric Broadcast Channel [4].
Let the input alphabet be $\mathcal{X} = \{0, 1\}$ and the output alphabets for recipients 1 and 2 be

$\mathcal{Y}_1 = \{0, 1\}$ and $\mathcal{Y}_2 = \{0, 1\}$, respectively. Let channel 1 be noiseless and channel 2 be a binary symmetric channel with cross-over probability p. The corresponding channel capacities are $C_1 = 1$ bit per transmission and $C_2(p) = 1 - h(p)$ bits per transmission, where $h(p)$ is the binary entropy function. The binary symmetric broadcast channel is a degraded broadcast channel since the condition (3.1) is satisfied.

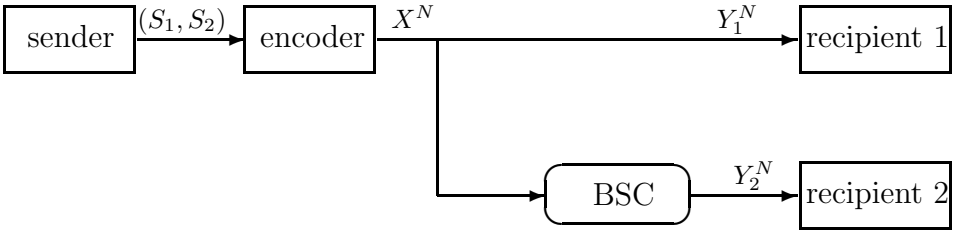

Figure 3.2: The binary symmetric broadcast channel.

The rate pair $(R_1, R_2) = (C_2(p), C_2(p))$ can be achieved using a standard $(2^{N[C_2(p)-\epsilon]}, N)$ code for channel 2 while the pair $(1, 0)$ can be achieved with no coding. By time-sharing the two strategies, the rate pairs on the straight line connecting the two points can be achieved, i.e., the points $(t_1C_1, t_2C_2(p))$ can be achieved for all $\{t_1, t_2\}$ such that $t_1, t_2 \geq 0$ and $t_1 + t_2 = 1$.

The code-partitioning strategy based on a good code selected from a randomly generated ensemble of codes and a minimum distance decoding rule can be used to enlarge the achievable rate region. The idea is to have two kinds of messages. A common message is communicated to both recipients 1 and 2 while an additional message is communicated to only recipient 1. In this way, the message intended for recipient 1 contains both the common message and the additional message whereas the message intended for recipient 2 contains only the common message.

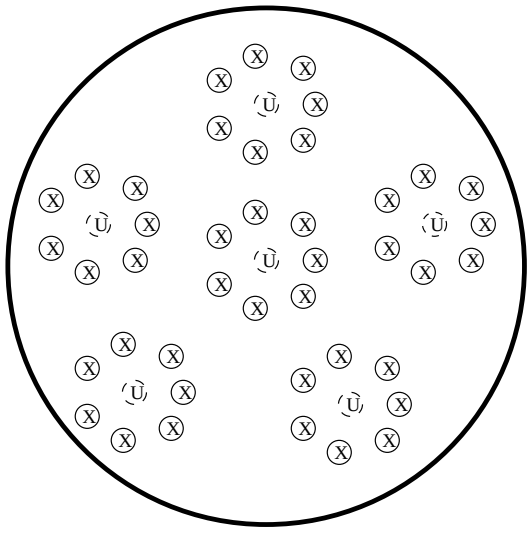

Figure 3.3: The codebook for the degraded broadcast channel

The strategy uses an auxiliary code C^{aux} and a satellite code C^{sat}. The auxiliary code is designed for a BSC with cross-over probability $p\bar{q} + \bar{p}q$ to be used at rate $C_2(p\bar{q} + \bar{p}q) - \epsilon$, where

$0 \leq p, q \leq 1$ $\bar{p} = 1 - p$, and $C_2(p\bar{q} + \bar{p}q)$ is the capacity of the channel. Based on the auxiliary code, the satellite code is designed so that all the vectors with Hamming distance qN from an auxiliary codeword are grouped into a partition, where N is the length of the codewords. There consequently are $2^{N[C_2(p\bar{q}+\bar{p}q)-\epsilon]}$ partitions in the satellite code, and each partition contains $\binom{N}{qN}$ satellite codewords. Each auxiliary codeword is uniquely assigned to a common message, and each satellite codeword in a partition is uniquely assigned to an additional message.

To send a pair of messages, use the common message to select an auxiliary codeword u^N, and use the additional message to select a satellite codeword x^N in the partition associated with the common message. Transmit the selected satellite codeword x^N through the broadcast channel. Recipient 2 receives the vector y_2^N and perceives it as the auxiliary codeword passing through 2 BSCs. The first one is the virtual BSC with cross-over probability q, resulting from the noise introduced by translation of the auxiliary codeword to the satellite codeword specified by the additional message. The second one is the BSC with cross-over probability p in the broadcast channel. Recipient 2 can reliably decode for the common message because the auxiliary code is designed for the cascaded channel. Recipient 1 receives the transmitted satellite codeword as y_1^N without error and therefore can decode for both the common message and the additional message.

Since the indices to the auxiliary codewords and the satellite codewords can be selected independently, the rates $R_1 = C_2(p\bar{q} + \bar{p}q) + \log[\binom{N}{qN}]/N - \epsilon \approx C_2(p\bar{q} + \bar{p}q) + h(q) - \epsilon$ and $R_2 = C_2(p\bar{q} + \bar{p}q) - \epsilon$ can be simultaneously achieved.

Gaussian Broadcast Channel [4].
The Gaussian variant of the broadcast channel is the case, in which channels 1 and 2 in the broadcast channel contain AWGN with means zero and variances N_1 and N_2, respectively. Without loss of generality, let $N_1 < N_2$. Let there be an average power constraint on the transmission power given by

$$\frac{1}{N}\sum_{i=1}^{N} x_i^2 \leq P,$$

where x^N is an input sequence. Channels 1 and 2 thus have the capacities of $C_1 = \frac{1}{2}\log(1 + P/N_1)$, and $C_2 = \frac{1}{2}\log(1 + P/N_2)$, respectively. The Gaussian broadcast channel is also a degraded broadcast channel. The rate pairs $(C_1, 0)$, (C_2, C_2) and the pairs resulting from time-sharing the two codes are achievable by the same arguments used in the binary symmetric broadcast channel.

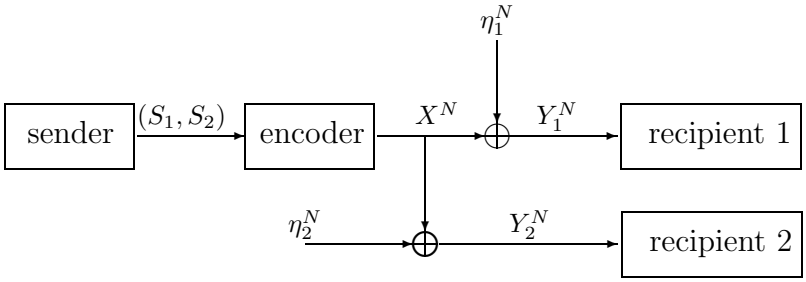

Figure 3.4: The Gaussian broadcast channel.

The achievable rate region of the Gaussian broadcast channel produced by the time-sharing strategy can also be enlarged by the use of the code-partitioning strategy as follows. Let's

distribute power $\bar{q}P$ to the transmission of the common message intended for both recipients 1 and 2 and qP to the transmission of the additional message intended for recipient 1; hence, recipient 2 perceives the power qP as additional noise power added to N_2, where q and $\bar{q}$ are constants to be specified such that $q + \bar{q} = 1$.

Let C^{aux} be an auxiliary code designed for an AWGN channel with average input power constraint $\bar{q}P$ and noise variance $qP + N_2$ at rate $C_2(\bar{q}P, qP + N_2) - \epsilon$, where $C_2(\bar{q}P, qP + N_2)$ is the capacity of the AWGN channel. Let C^{sat} be a satellite code designed for the second AWGN channel with average input power constraint qP and noise variance N_1 at rate $C_1(qP, N_1) - \epsilon$, where $C_1(qP, N_1)$ is the capacity of the second AWGN channel.

To send a pair of messages, use the common message to select an auxiliary codeword u^N in C^{aux}, and use the additional message to select a satellite codeword x^N in C^{sat}. Transmit $u^N + x^N$ through the broadcast channel. Recipient 2 receives the vector y_2^N and perceives it as the auxiliary codeword passing through two AWGN channels. The first one is the virtual AWGN channel with noise power qP, resulting from the noise introduced by the addition of x^N to u^N. The second one is the AWGN channel with noise power N_2.

Recipient 2 can reliably decode the received vector y_2^N for u^N and the common message because the auxiliary code is designed for the cascaded channel. Similarly, recipient 1 can reliably decode the received vector y_1^N for u^N because $N_1 < N_2$ and subsequently decode the vector $y_1^N - u^N$ for x^N and the additional message. Hence, recipient 2 can reliably decode for the common message at rate

$$R_2 = \frac{1}{2} \log \left[1 + \frac{\bar{q}P}{qP + N_2} \right] - \epsilon.$$

Recipient 1 can reliable decode for both the common and the addition messages at rate

$$R_1 = \frac{1}{2} \log \left[1 + \frac{\bar{q}P}{qP + N_2} \right] + \frac{1}{2} \log \left[1 + \frac{qP}{N_1} \right] - 2\epsilon.$$

Varying the power distribution factor q yields the rate pairs dominating the time-sharing rate pairs.

The capacity region for the degraded broadcast channel, proved by Bergmans in [1] and [2] by Gallager in [11], coincides with the regions in the above two cases. The capacity region of the general broadcast channel is still unknown.

3.3 Wiretap Channel

Communication through a wiretap channel is another interesting situation, in which the code-partitioning strategy can be used, though in a different way for a different purpose from those in the broadcast channel. The wiretap channel was introduced and investigated by Wyner in [20]. In this communication model, a message is communicated through a discrete memoryless main channel subjected to a wiretapper observing the output of the main channel through another discrete memoryless wiretap channel. A system designer seeks to encode the message so that it can be transmitted reliably at a high rate to a recipient and is kept secret from the

wiretapper. This model can be thought of as a type of broadcast channel with a different design objective of maximizing the rate to the recipient and minimizing the rate to the wiretapper, as opposed to maximizing the two rates. The secrecy of the message is measured by using the concept of equivocation.

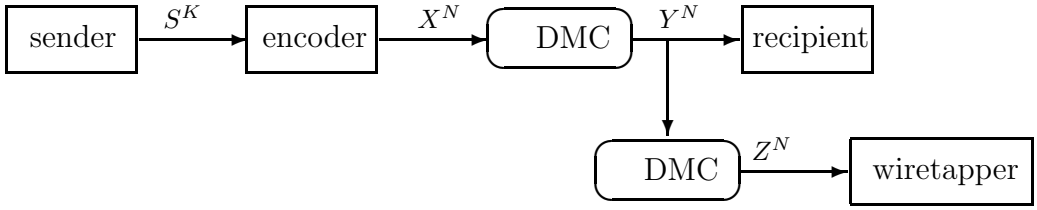

Figure 3.5: The wiretap channel.

Let $S^K \in \mathcal{S}^K$ be a sequence of i.i.d. random variables representing a message with entropy $H(S^K) = KH(S)$ to be encoded as $X^N \in \mathcal{X}^N$, which is the input to the main channel. $Y^N \in \mathcal{Y}^N$ is the output of the main channel received by the intended recipient and is observed by the wiretapper through the wiretap channel as $Z^N \in \mathcal{Z}^N$, where $\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are finite sets. Let $\hat{s}^K \in \mathcal{S}^K$ be the message decoded by the recipient from Y^N. The rate of communication to the recipient is defined as $H(S^K)/N$, and the equivocation to the wiretapper is defined as $\Delta = H(S^K|Z^N)/K$. When the equivocation is $H(S^K)/K$, the perfect secrecy is achieved, which means that S^K and Z^N are independent, and the wiretapper is not better off in detecting the message S^K given the observation Z^N. A rate-equivocation pair (R, d) is said to be achievable if, for all $\epsilon > 0$, there exists an encoder-decoder pair for which

$$\begin{aligned} \frac{H(S^K)}{N} &\geq R - \epsilon \\ \frac{H(S^K|Z^N)}{K} &\geq d - \epsilon \\ P_e &\leq \epsilon, \end{aligned}$$

where

$$P_e = \Pr\{\hat{S}^K \neq S^K\}.$$

Denoting the capacity of the main channel by $C_M = \sup_{p_X} I(X;Y)$ and the capacity of the combined (main-wiretap) channel by $C_{MW} = \sup_{p_X} I(X;Z)$, Wyner proved that the capacity region is

$$\bar{\mathcal{R}} = \left\{(R,d) : 0 \leq R \leq C_M, 0 \leq d \leq \frac{H(S^K)}{K}, Rd \leq \frac{H(S^K)}{K}\Gamma(R)\right\},$$

where

$$\begin{aligned} \Gamma(R) &= \sup_{p_X \in \mathcal{P}(R)} I(X;Y|Z), \\ \mathcal{P}(R) &= \{p_X : I(X;Y) \geq R \text{ for } R \geq 0\}. \end{aligned}$$

The coding strategy that is used to prove the region is illustrated for the rate-equivocation pair $(C_s, H(S^K)/K)$ here, where C_s is the maximum achievable rate in perfect secrecy =

$\max_{(R,H(S^K)/K)\in\bar{\mathcal{R}}} R$. To begin, design a code C^1 that achieves the capacity of the main channel so that there are $2^{N[C_M-\epsilon]}$ codewords for a given $\epsilon > 0$. Then partition the code C^1 into subcodes so that each subcode can be used for communication through the combined channel at rate $C_{MW} - \epsilon$. There are $2^{N[C_s-\epsilon]}$ subcodes, where $C_s = C_M - C_{MW}$.

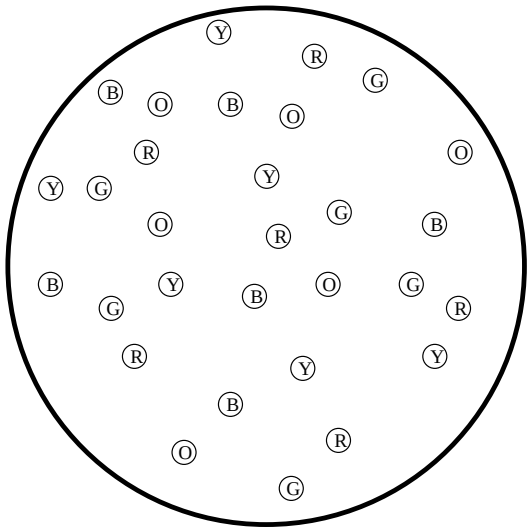

Figure 3.6: Coloring the codewords. R,G,B,O,Y represent the colors red, green, blue, orange and yellow, respectively.

We can uniquely assign a color to each subcode so that all the codewords in the same subcode has the same color. Let each color represent a message for the intended recipient. To send a message, the encoder randomly selects one of $2^{N[C_{MW}-\epsilon]}$ codewords in the subcode associated with the message to be transmitted. The recipient can reliably detect the codeword; thereby identifying the color or the codeword and the message. The wiretapper can also reliably detect the transmitted codeword provided the knowledge of the subcode because each subcode is a code for the combined channel. However, the subcode being used cannot then be identified by the wiretapper for all the subcodes are possible given the wiretapper's observation resulting in the perfect secrecy of the message. An example of the code based on this coding strategy is given for a special case of the wiretap channel in the next section.

3.4 Wiretap Channel II

The wiretap channel II is a special case of the wiretap channel with a more powerful wiretapper in that the main channel is noiseless ($X^N = Y^N$), and the wiretapper can examine a size-μ subset of the codeword components of his choice. For the choice of a subset τ^c of the components of a codeword, $Z^\mu = X^N_{\tau^c}$, where $\tau^c \subseteq \{1, 2, \ldots, N\}, |\tau^c| = \mu$ and $X^N_{\tau^c}$ is the length-μ vector containing components of X^N indexed by the members of τ^c, and members of τ^c are always in an increasing order.

Ozarow and Wyner examined this situation, in which all the random variables have binary

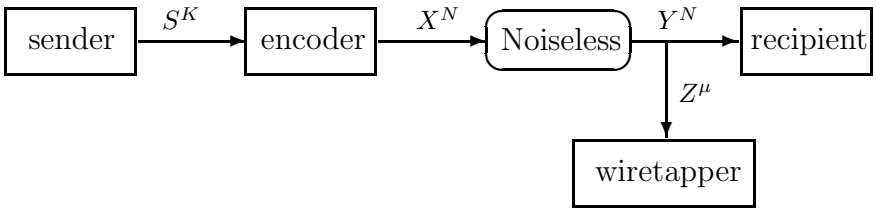

Figure 3.7: The wiretap channel II

alphabets in [16] by redefining the equivocation to be

$$\Delta \quad = \quad \min_{\tau^c:|\tau^c|=\mu} H(S^K|Z^\mu).$$

The minimization in the definition of the equivocation reflects the best that the wiretapper can do. A system designer tries to find a coding scheme that provides an optimal tradeoff between the rate of transmission and the equivocation. In this case, the code-partitioning strategy is used together with group codes to prove the rate-equivocation capacity region for the channel given the wiretapper's capability parameter μ.

The coding strategy makes use of a $K \times N$ binary parity-check matrix $\mathbf{A}_{K\times N}$ to send a K-bit message using an N-bit codeword. To send a message s^K, the encoder solves the simultaneous equations $\mathbf{s}^T = \mathbf{A}\mathbf{x}^T$ for a set of possible $\mathbf{x}^T$ and randomly selects one of the solutions to be transmitted, where $\mathbf{s}^T = (s^K)^T$, and $\mathbf{x}^T = (x^N)^T$. The intended recipient calculates s^K from y^N and $\mathbf{A}$ without error because the main channel is noiseless. The set of solutions to $\mathbf{s}^T = \mathbf{A}\mathbf{x}^T$ is a partition of length-N vectors in the N-dimensional space corresponding to the message s^K and a color as described in the previous section.

We now illustrate the partitioning of a code for communicating 2-bit information using 3-bit codewords. Let a code $C^1 = \{000, 001, 010, 011, 100, 101, 110, 111\}$, and colors red, green, blue, and yellow correspond to information 00, 01, 10 and 11, respectively. Let's partition the code C^1 by using the matrix $\mathbf{A}$ so that the subcode corresponding to a message s_1^K is the set of solution to $\mathbf{s}_1^T = \mathbf{A}\mathbf{x}^T$. For

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

and the corresponding partition is

color	message	codewords
red	00	000, 111
green	01	010, 101
blue	10	011, 100
yellow	11	001, 110

Hence, there are four subcodes. The red, green, blue and yellow subcodes respectively are $\{000, 111\}$, $\{010, 101\}$, $\{011, 100\}$, $\{001, 110\}$. When the message to be sent is 11, the encoder randomly selects either 001 or 110 to be transmitted over the noiseless channel, and the recipient correctly decodes for the message. If the wiretapper observes the first two positions of the transmitted codeword, there are two possible messages (00 and 11) corresponding to 1 bit of

uncertainty about the message. A theoretical analysis of the uncertainty is given in Chapter 6.

Given the parameter μ, a triple (R, α, d) is achievable if, for all $\epsilon > 0$ and all integers $N_0 > 0$, there exists an encoder-decoder pair with parameters $N \geq N_0$, $K \geq (R - \epsilon)N$, $\mu \geq (\alpha - \epsilon)N$, $\Delta \geq (d - \epsilon)K$, and $P_e \leq \epsilon$. The capacity region was given by the set of (R, α, d) such that $0 \leq R, \alpha \leq 1$, and

$$0 \leq d \leq \begin{cases} 1, & \text{for } 0 \leq \alpha \leq 1 - R, \\ \frac{1-\alpha}{R}, & \text{for } 1 - R \leq \alpha \leq 1. \end{cases}$$

3.5 Gaussian Wiretap Channel

The Gaussian wiretap channel is another variant of the wiretap channel. The main and the wiretap channels are AWGN channels with i.i.d. noise distributions $\eta_1 \sim \mathcal{N}(0, \sigma_{\eta_1}^2)$ and $\eta_2 \sim \mathcal{N}(0, \sigma_{\eta_2}^2)$, respectively. The message $S^K \in \mathcal{S}^K$ has finite alphabet. The input X^N to the main

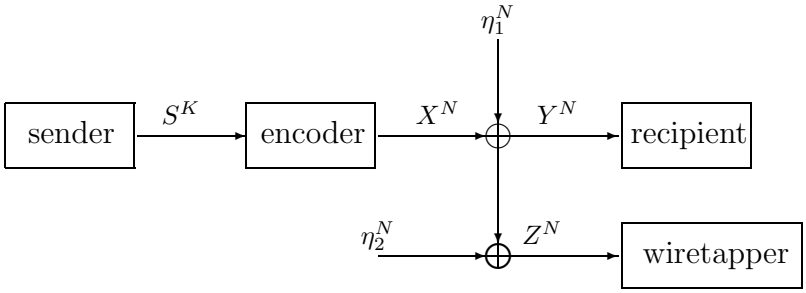

Figure 3.8: The Gaussian wiretap channel.

channel has continuous alphabet with an average power constraint:

$$\frac{1}{N} \sum_{i=1}^{N} E[X_i^2] \leq P.$$

The capacities of the main channel and the combined channel C_M and C_{MW} respectively are

$$\begin{aligned} C_M &= \frac{1}{2} \log \left[1 + \frac{P}{\sigma_{\eta_1}^2} \right] \\ C_{MW} &= \frac{1}{2} \log \left[1 + \frac{P}{\sigma_{\eta_1}^2 + \sigma_{\eta_2}^2} \right]. \end{aligned}$$

Leung-Yan-Cheong and Hellman modified the definition of the equivocation so that $\Delta = H(S^K|Z^N)/H(S^K)$ and proved the capacity region of the Gaussian wiretap channel in [14]. The achievability part was proved by time-sharing two codes respectively achieving the rate-equivocation pairs $(C_M, C_s/C_M)$ and $(C_s, 1)$, where $C_s = C_M - C_{MW}$. The first code is designed to achieve the capacity of the main channel as if the wiretapper is absent. The second code is selected from a code ensemble with the following properties to ensure that almost all

codes in the ensemble achieve the pair $(C_s, 1)$ and satisfy the average power constraint, where $\epsilon > 0$:

1. each code in the ensemble has message rate $C_s - \epsilon$;
2. components of each codeword are i.i.d. random variables with distribution $\mathcal{N}(0, P-\alpha)$, where $\alpha > 0$ is chosen so that
   - $C_M(\alpha) = \frac{1}{2}\log(1 + (P-\alpha)/\sigma_{\eta_1}^2) > C_M - \epsilon$;
   - $C_{MW}(\alpha) = \frac{1}{2}\log(1 + (P-\alpha)/(\sigma_{\eta_1}^2 + \sigma_{\eta_2}^2)) > C_{MW} - \epsilon$.

The capacity of the channel was proved to be the set of rate-equivocation pairs (R, d) satisfying the conditions:

$$\begin{aligned} R &\leq C_M \\ d &\leq 1 \\ Rd &\leq C_s. \end{aligned}$$

3.6 Dirty-paper Channel

Unlike the broadcast channels and the wiretap channels, the dirty-paper channel usually involves one sender and one recipient without any wiretapper. The sender wishes to send a message to a recipient through an AWGN channel with i.i.d. noise $\eta_1 \sim \mathcal{N}(0, \sigma_{\eta_1}^2)$ and additive i.i.d. interference $V \sim \mathcal{N}(0, \sigma_V^2)$. With the complete knowledge of the interference sequence V^N in advance, the sender seeks to encode his message into an appropriate codeword for the channel. This is a special case of the channel with random parameters investigated in [12] by Gel'fand and Pinsker.

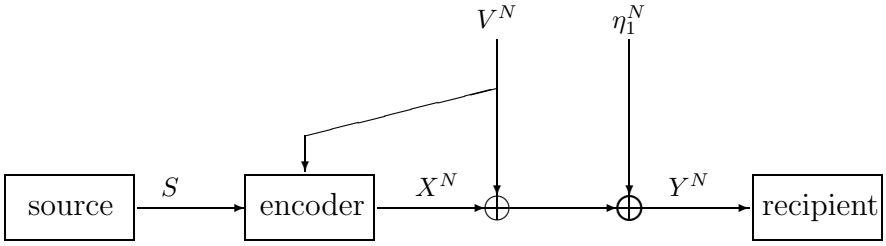

Figure 3.9: The dirty-paper channel.

In [3], Costa proposed a coding scheme that compensates for the interference so that it, when known in advance to the encoder, does not degrade the rate of transmission. The capacity of the dirty-paper channel is

$$C_{dirty} = \frac{1}{2}\log\left[1 + \frac{P}{\sigma_{\eta_1}^2}\right].$$

The capacity is achieved by making use of an auxiliary random variable $U \sim \mathcal{N}(0, P + \alpha^2\sigma_V^2)$, where α is a real number. The codebook in the coding scheme consists of $2^{N[C_{dirty}-\epsilon]}$ bins,

each of which contains $2^{N[I(U;V)+\epsilon/2]}$ auxiliary codewords u^N generated i.i.d. from U. Each bin is associated with a message. To send a message s^K, the encoder looks for an auxiliary codeword in the bin associated with the message that is jointly typical with the interference v^N. The corresponding codeword is calculated as $x^N = u^N - \alpha v^N$ to be transmitted. The recipient observes $y^N = x^N + v^N + \eta_1^N$ and looks for an auxiliary codeword u^N that is jointly typical with y^N. The recipient declares the bin containing the auxiliary codeword u^N as the detected message. For sufficiently large N, the message can be communicated reliably at the rate close to the capacity.

In this situation, the auxiliary codeword is used to convey the message and sufficient information about the interference to the recipient so that he can detect the message reliably. The binning in this codebook is similar to the color assignment in the codebook for the wiretap channel: that is each bin corresponds to a color (message); however, the selection of codeword in this situation depends on the interference as opposed to being random in the wiretap channel.

3.7 Remarks

Even though the broadcast, wiretap and dirty-paper situations posses different communication objectives, the code-partitioning strategy can be used in different ways to attain the goals. The purposes and the usages of code partitioning are tabulated in Table 3.1. This coding strategy will be used to prove an achievable rate-equivocation region for the Gaussian wiretap channel with side information in the next chapter.

Channel	Usage of code partitioning	Codeword
Broadcast	• An auxiliary codeword representing the center of a cloud is used as an index to a code partition. • The partition index is a common message to both recipient 1 and recipient 2. • The distance between a satellite codeword and its cloud center is an additional noise for recipient 2.	• Each codeword is a satellite associated with a center of a cloud. • The satellite codewords associated with the same cloud tend to stay close together.
Wiretap	• A subcode (partition) index is associated with a color. • Each partition index represents a message to the recipient.	• The codewords, associated with the message, not being transmitted are used to confuse the wiretapper. • The codewords associated with the same message tend to spread out.
Dirty Paper	• A subcode (partition) index is associated with a color. • Each partition index represents a message to the recipient.	• The codewords associated with the message being transmitted are used to convey information about the interference to the recipient. • The codewords associated with the same message tends to spread out.

Table 3.1: Comparison of the code-partitioning uses for different channels.

Chapter 4

Gaussian Wiretap Channel with Side Information

4.1 Introduction

The Gaussian wiretap channel (GWC) is extended to the Gaussian wiretap channel with side information (GWCSI) in this chapter. For the analysis of the two models, we introduce leakage functions and propose three modes of operation of a perfect-secrecy coding strategy based on the code-partitioning concept. The interference-free mode is used for the GWC and used as the foundation for the camouflage and the high-power modes developed for the GWCSI. The camouflage and the high-power modes play crucial roles in the proof of an achievable region for the GWCSI.

4.2 Gaussian Wiretap Channel

Although the capacity region of the GWC was proved in [14], a specific encoding-decoding technique was not provided. In this section, we introduce the interference-free mode of operation, which includes codebook generating, encoding and decoding processes.

4.2.1 Model Description

Let S^K be a finite-alphabet message and X_c be the codeword generated from S^K by an encoding process with an average power constraint P. Let $Y^N = X_c + \eta_1^N$, and $Z^N = Y^N + \eta_2^N$ be outputs of the main channel and the wiretap channel, where η_1^N and η_2^N are sequences of independent random variables identically distributed according to $\mathcal{N}(0, N_1)$ and $\mathcal{N}(0, N_2)$, respectively. Hence, the capacities of the main channel and the combined channel respectively

are:

$$\begin{aligned} C_M &= \frac{1}{2}\log\left[\frac{P+N_1}{N_1}\right] \\ C_{MW} &= \frac{1}{2}\log\left[\frac{P+N_1+N_2}{N_1+N_2}\right], \end{aligned}$$

and the secret capacity is defined as

$$C_s = C_M - C_{MW} = \frac{1}{2}\log\left[\frac{(P+N_1)(N_1+N_2)}{N_1(P+N_1+N_2)}\right].$$

The recipient decodes the output of the main channel Y^N for $\hat{S}^K$ at rate K/N with probability of error

$$P_e = \Pr\{\hat{S}^K \neq S^K\}$$

and equivocation $H(S^K|Z^N)/H(S^K)$. We say that the rate-equivocation pair (R, d) is achievable if, for a given $\epsilon > 0$, there exists an encoder-decoder pair such that

$$\begin{aligned} \frac{H(S^K)}{N} &\geq R-\epsilon \\ \frac{H(S^K|Z^N)}{H(S^K)} &\geq d-\epsilon \\ P_e &\leq \epsilon. \end{aligned}$$

4.2.2 Leakage Function

The mutual information $I(X;Y)$ is the average information rate from an input X to an output Y through a channel $p(Y|X)$. In the GWC, there are one input X and two outputs: Y at the recipient and Z at the wiretapper. $I(X;Y)$ is the average information between the encoder and the recipient while $I(X;Z)$ is that between the encoder and the wiretapper. Hence, $I(X;Z)$ can be defined as the leakage function for the GWC, which corresponds to average information leakage rate from the main channel through the wiretap channel.

4.2.3 Interference-free Mode

The design of the interference-free mode of operation is based on the code-partitioning concept introduced in Chapter 3, the leakage function and the jointly typical decoder. The interference-free mode is defined by codebook generating, encoding and decoding processes. The main idea is to partition a code designed for the main channel so that each partition is a code for the combined (main-wiretap) channel, which can be used to communicate at the average information leakage rate.

To describe the interference-free mode, we make use of random variables $\tilde{X}$, $\tilde{Y}$ and $\tilde{Z}$ to determine the sizes of the codebook and the partitions. Now, the random variables involved are

defined as follows:

$\tilde{X} \sim \mathcal{N}(0, P)$, $\tilde{Y} \sim \mathcal{N}(0, P + N_1)$, $\tilde{Z} \sim \mathcal{N}(0, P + N_1 + N_2)$
$X \sim \mathcal{N}(0, P')$, $Y \sim \mathcal{N}(0, P' + N_1)$, $Z \sim \mathcal{N}(0, P' + N_1 + N_2)$
X_c represents the output of the codeword selection process,
where $P' = P(1 + 2\delta \ln(2))^{-1} < P$ and $\delta > 0$.

Let $\tilde{X}^N$, $\tilde{Y}^N$, $\tilde{Z}^N$ be sequences of i.i.d. Gaussian random variables $\tilde{X}$, $\tilde{Y}$ and $\tilde{Z}$, respectively. Given an $\epsilon > 0$, let δ be a constant associated with the encoding and decoding processes, and ξ be a constant associated with the wiretapper's decoding process to be specified later. Then, the interference-free mode can be described as follows:

1. *Generating the codebook.* Generate $2^{N[I(\tilde{X};\tilde{Y})-2\epsilon]}$ sequences x^N according to the distribution $p(x^N) = \prod_{i=1}^{N} p(x_i)$, and $p(x_i) \sim \mathcal{N}(0, P')$ for all $i \in \{1, 2, \ldots, N\}$. Place the sequences x^N randomly into bins so that each bin contains $2^{N[I(\tilde{X};\tilde{Z})-\epsilon]}$ sequences, creating $2^{N[I(\tilde{X};\tilde{Y})-I(\tilde{X};\tilde{Z})-\epsilon]}$ bins. Index each bin by $j \in \{1, 2, \ldots, 2^{NR}\}$, where $R = I(\tilde{X};\tilde{Y}) - I(\tilde{X};\tilde{Z}) - \epsilon$. The codebook is given to the sender and the recipient.

2. *Encoding.* To send a message j, the sender randomly selects a sequence x_c in bin j such that $x_c \in T_X^N(\delta)$ to be transmitted.

3. *Decoding.* To decode the received sequence for the message, the recipient finds a sequence x_c in the codebook that is jointly typical with the received sequence y^N, i.e., $(x_c, y^N) \in T_{X,Y}^N(\delta)$. Declare the index to the bin, in which the sequence is found as the received message.

4. *Wiretapper's decoding.* The wiretapper receives a sequence z^N and finds a sequence x_c in the codebook that is jointly typical with the received sequence, i.e., $(x_c, z^N) \in T_{X,Z}^N(\xi)$. Declare the index to the bin, in which the sequence is found as the received message.

5. *Probability of error.* An error occurs in the coding process when a message j is to be transmitted and one or more of the following events occurs.

   - $\mathcal{E}^X(j)$: in the encoding process, there is no sequence x_c in bin j that satisfies the power constraint.
   - $\mathcal{E}^{Y1}(j)$: in the decoding process, there is no sequence x_c in the codebook that is jointly typical with the received sequence.
   - $\mathcal{E}^{Y2}(j)$: in the decoding process, a sequence x_c in bin $i \neq j$ is jointly typical with the received sequence.

Note that this mode of operation is an analogue of the coding strategy used in the wiretap channel for the Gaussian wiretap channel . In Theorem 4.1, we prove the achievability of the rate-equivocation pair $(C_s, 1)$ for the GWC using the interference-free mode.

Theorem 4.1 *Given the interference-free mode to be used in the Gaussian wiretap channel with average power constraint P, for any $\epsilon > 0$, there exists an integer N_{min} such that, for*

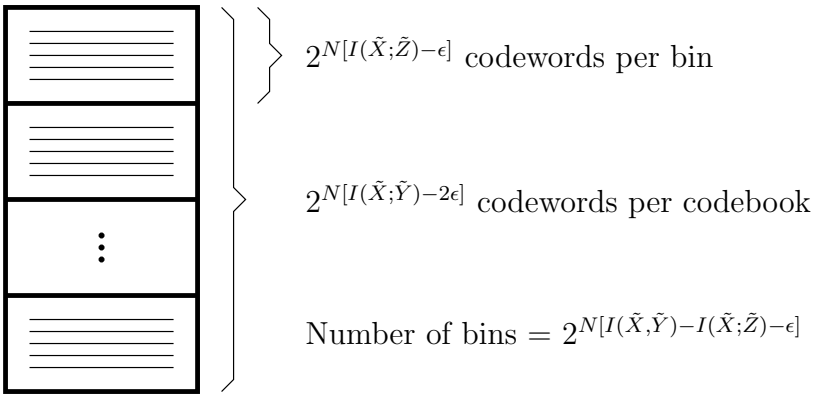

Figure 4.1: The codebook in the interference-free mode for the Gaussian wiretap channel.

$N \geq N_{min}$,

$$\begin{aligned} \frac{H(S^K)}{N} &\geq \frac{1}{2}\log\left[\frac{(P+N_1)(N_1+N_2)}{N_1(P+N_1+N_2)}\right]-\epsilon \\ \frac{H(S^K|Z^N)}{H(S^K)} &\geq 1-\epsilon \\ P_e &\leq \epsilon. \end{aligned}$$

Proof: This proof consists of three main parts: proof of the rate, proof of the probability of error, and proof of the equivocation.

Proof of the Rate. Given $\epsilon > 0$, the number of messages that can be sent is the number of bins in the codebook: $2^{H(S^K)} = 2^{NR} = 2^{N[I(\tilde{X};\tilde{Y})-I(\tilde{X};\tilde{Z})-\epsilon]}$. Hence, by the design of the codebook,

$$\begin{aligned} R \triangleq \frac{H(S^K)}{N} &= I(\tilde{X};\tilde{Y}) - I(\tilde{X};\tilde{Z}) - \epsilon \\ &= \frac{1}{2}\log\left[\frac{P+N_1}{N_1}\right] - \frac{1}{2}\log\left[\frac{P+N_1+N_2}{N_1+N_2}\right] - \epsilon \\ &= \frac{1}{2}\log\left[\frac{(P+N_1)(N_1+N_2)}{N_1(P+N_1+N_2)}\right] - \epsilon. \end{aligned}$$

Proof of the probability of error. The probability of error is bounded by the probabilities of the events: $\{\mathcal{E}^X(1)\}$, $\{\mathcal{E}^{Y1}(1)|\mathcal{E}^X(1)^C\}$, and $\{\mathcal{E}^{Y2}(1)|\mathcal{E}^X(1)^C\}$, each of which can be bounded by $\epsilon/3$ in the following steps.

Let's denote the k^{th} codeword in bin j by $x_c(j,k)$, where $k \in \{1,2,\ldots,2^{N[I(\tilde{X};\tilde{Z})-\epsilon]}\}$ and the randomly selected codeword to be transmitted, given a message j, by $x_c(j,k^*)$.

$$\begin{aligned} P_e &\overset{(1)}{\leq} \sum_{j=1}^{2^{NR}} \Pr\{S^K=j\}[\Pr\{\mathcal{E}^X(j)\} + \Pr\{\mathcal{E}^{Y1}(j)|\mathcal{E}^X(j)^C\} + \Pr\{\mathcal{E}^{Y2}(j)|\mathcal{E}^X(j)^C\}] \\ &\overset{(2)}{=} \Pr\{\mathcal{E}^X(1)\} + \Pr\{\mathcal{E}^{Y1}(1)|\mathcal{E}^X(1)^C\} + \Pr\{\mathcal{E}^{Y2}(1)|\mathcal{E}^X(1)^C\}. \end{aligned}$$

(1) follows from the union bound (See Appendex A). (2) follows from the fact that error events don't depend on the message j. We now consider the three terms in the summation separately.

$\Pr\{\mathcal{E}^X(1)\}$. By Lemma 2.3 with $\epsilon_1 = \epsilon/3$ and a given $\delta > 0$, there exists an integer L_0 such that $\Pr\{X_c \in T_X^N(\delta)\} > 1 - \epsilon/3$ for $N \geq L_0$, which implies that $\Pr\{\mathcal{E}^X(1)\} \leq \epsilon/3$ for given $\epsilon > 0$, $\delta > 0$ and $N \geq L_0$ based on Lemma 2.9.

$\Pr\{\mathcal{E}^{Y1}(1)|\mathcal{E}^X(1)^C\}$. By Lemma 2.5 with $\epsilon_1 = \epsilon/3$ and a given $\delta > 0$, there exists an integer L_1 such that $\Pr\{(X_c, Y^N) \in T_{X,Y}^N(\delta)\} > 1 - \epsilon/3$ for $N \geq L_1$, which implies that $\Pr\{\mathcal{E}^{Y1}(1)|\mathcal{E}^X(1)^C\} \leq \epsilon/3$ for given $\epsilon > 0$, $\delta > 0$ and $N \geq L_1$.

$\Pr\{\mathcal{E}^{Y2}(1)|\mathcal{E}^X(1)^C\}$. By the code generating process, $X_c(1, k^*)$ and $X_c(i, k)$ are independent, and so are Y^N and $X_c(i, k)$ for $(i, k) \neq (1, k^*)$. By inequality (2.20), $\Pr\{(X_c(i, k), Y^N) \in T_{X,Y}^N(\delta)\} < 2^{-N[I(X;Y)-3\delta]}$ for $(i, k) \neq (1, k^*)$. Hence,

$$\begin{aligned}
\Pr\{\mathcal{E}^{Y2}(1)|\mathcal{E}^X(1)^C\} &\overset{(3)}{<} \sum_{(i,k)\neq(1,k^*)} 2^{-N[I(X;Y)-3\delta]} \\
&= (2^{N[I(\tilde{X};\tilde{Y})-2\epsilon]} - 1)2^{-N[I(X;Y)-3\delta]} \\
&< 2^{-N[I(X;Y)-3\delta-I(\tilde{X};\tilde{Y})+2\epsilon]} \\
&\leq \epsilon/3
\end{aligned}$$

when

$$I(X;Y) - I(\tilde{X};\tilde{Y}) - 3\delta + 2\epsilon > 0,$$

and

$$\begin{aligned}
N[I(X;Y) - I(\tilde{X};\tilde{Y}) - 3\delta + 2\epsilon] &\geq \log\left[\frac{3}{\epsilon}\right] \\
N &\geq L_2 = \frac{1}{I(X;Y) - I(\tilde{X};\tilde{Y}) - 3\delta + 2\epsilon} \log\left[\frac{3}{\epsilon}\right].
\end{aligned}$$

(3) follows from the union bound.

The two conditions above impose restrictions on δ and N, respectively. To obtain the restric-

tions, define $g_{LHS}(\delta) = I(X;Y) - I(\tilde{X};\tilde{Y})$, and $g_{RHS}(\delta) = 3\delta - 2\epsilon$. Then,

$$\begin{aligned}
g_{LHS}(\delta) &\overset{(4)}{=} \frac{1}{2}\log\left[\frac{P(1+2\delta\ln(2))^{-1}+N_1}{N_1}\right] - \frac{1}{2}\log\left[\frac{P+N_1}{N_1}\right] \\
&= \frac{1}{2}\log\left[\frac{P(1+2\delta\ln(2))^{-1}+N_1}{P+N_1}\right] \\
g_{LHS}(0) &= 0 \\
\lim_{\delta\to\infty} g_{LHS}(\delta) &= \frac{1}{2}\log\left[\frac{N_1}{P+N_1}\right] < 0 \\
\frac{dg_{LHS}(\delta)}{d\delta} &= \left[\frac{1}{2\ln(2)}\right]\left[\frac{P+N_1}{P(1+2\delta\ln(2))^{-1}+N_1}\right]\left[\frac{-P}{P+N_1}\right](1+2\delta\ln(2))^{-2}2\ln(2) \\
&= \frac{-P(1+2\delta\ln(2))^{-2}}{P(1+2\delta\ln(2))^{-1}+N_1} \\
&= \frac{-P(1+2\delta\ln(2))^{-1}}{P+N_1(1+2\delta\ln(2))} \\
&< 0 \text{ because } P, N_1, \delta > 0.
\end{aligned}$$

(4) follows from equation (2.13). From the facts that $g_{LHS}(0) = 0$, $g_{RHS}(0) = -2\epsilon < 0$, and $g_{LHS}(\delta)$ is a monotonically decreasing continuous function converging to a negative number while $g_{RHS}(\delta)$ is a monotonically increasing continuous function approaching ∞ as δ increases, it follows that there exists a $\delta_0 > 0$ such that $g_{LHS}(\delta_0) = g_{RHS}(\delta_0)$ by the continuity of the functions [13]. See an example in Figure 4.2.

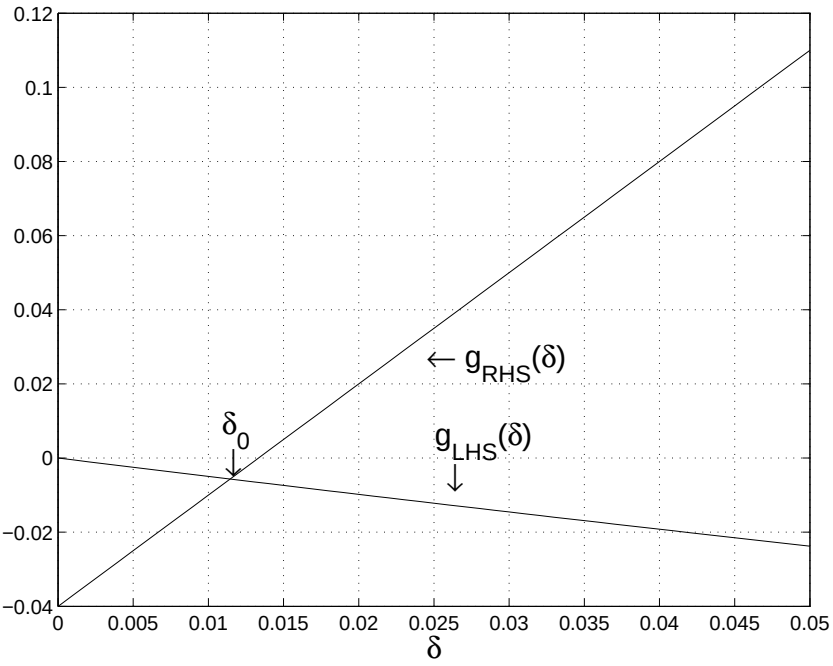

Figure 4.2: Calculating δ_0 for $P = 1, N_1 = 1$ and $\epsilon = 0.02$.

Thus, $g_{LHS}(\delta) - g_{RHS}(\delta) = I(X;Y) - I(\tilde{X};\tilde{Y}) - 3\delta + 2\epsilon > 0$ for $0 < \delta < \delta_0$, and $\Pr\{\mathcal{E}^{Y2}(1)|\mathcal{E}^{X}(1)^{C}\}$

$\leq \epsilon/3$ for $\delta < \delta_0$, and $N \geq L_2$, and

$$P_e \quad \leq \quad \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon$$

for $\delta < \delta_0$, and $N \geq \max\{L_0, L_1, L_2\}$.

Proof of the equivocation. This part of the proof is done in three steps:

1. show that $H(S^K|Z^N) = N[I(X;Y) - I(X;Z)] - H(X_c|S^K, Z^N)$;
2. show that $I(X;Y) - I(X;Z) \geq R(1-\epsilon/2)$;
3. show that $H(X_c|S^K, Z^N)/RN \leq \epsilon/2$.

Combining the above three steps,

$$\begin{aligned}
\frac{H(S^K|Z^N)}{H(S^K)} &= \frac{N[I(X;Y) - I(X;Z)] - H(X_c|S^K, Z^N)}{RN} \\
&\geq 1 - \epsilon/2 - \frac{H(X_c|S^K, Z^N)}{RN} \\
&\geq 1 - \epsilon/2 - \epsilon/2 \\
&= 1 - \epsilon.
\end{aligned}$$

Proceeding to the first step,

$$\begin{aligned}
H(S^K|Z^N) &\overset{(5)}{=} H(S^K, Z^N) - H(Z^N) \\
&\overset{(6)}{=} H(S^K, X_c, Z^N) - H(X_c|S^K, Z^N) - H(Z^N) \\
&\overset{(7)}{=} H(X_c, Z^N) + H(S^K|X_c, Z^N) - H(X_c|S^K, Z^N) - H(Z^N) \\
&\overset{(8)}{=} H(X_c|Z^N) - H(X_c|S^K, Z^N) \\
&\overset{(9)}{\geq} H(X_c|Z^N) - H(X_c|Y^N) - H(X_c|S^K, Z^N) \\
&= [H(X_c|Z^N) - H(X_c)] + [H(X_c) - H(X_c|Y^N)] - H(X_c|S^K, Z^N) \\
&\overset{(10)}{=} I(X_c; Y^N) - I(X_c; Z^N) - H(X_c|S^K, Z^N) \\
&\overset{(11)}{=} [H(Y^N) - H(Y^N|X_c)] - [H(Z^N) - H(Z^N|X_c)] - H(X_c|S^K, Z^N) \\
&\overset{(12)}{=} [H(Y^N) - H(\eta_1^N)] - [H(Z^N) - H(\eta_1^N + \eta_2^N)] - H(X_c|S^K, Z^N) \\
&\overset{(13)}{=} I(X^N; Y^N) - I(X^N; Z^N) - H(X_c|S^K, Z^N) \\
&\overset{(14)}{=} N[I(X;Y) - I(X;Z)] - H(X_c|S^K, Z^N),
\end{aligned}$$

where (5), (6) and (7) follow from identities (2.1), (2.2), and (2.3). (8) follows from the fact that the secret message s^K is known given a codeword x_c. (9) follows from the fact that $H(X_c|Y^N) \geq 0$. (10), (11) and (13) follows from equations (2.4) and (2.5). (12) follows from the fact that X_c, η_1^N, and η_2^N are independent. (14) follows from Lemma 2.1.

In step 2, apply equation (2.13) to $I(X;Y)$ and $I(X;Z)$ to obtain

$$\begin{aligned} I(X;Y) - I(X;Z) &= \frac{1}{2}\log\left[\frac{P' + N_1}{N_1}\right] - \frac{1}{2}\log\left[\frac{P' + N_1 + N_2}{N_1 + N_2}\right] \\ &= \frac{1}{2}\log\left[\frac{(P' + N_1)(N_1 + N_2)}{N_1(P' + N_1 + N_2)}\right]. \end{aligned}$$

Hence,

$$\begin{aligned} \frac{1}{2}\log\left[\frac{(P' + N_1)(N_1 + N_2)}{N_1(P' + N_1 + N_2)}\right] &\geq R(1 - \epsilon/2) \\ \frac{(P' + N_1)(N_1 + N_2)}{N_1(P' + N_1 + N_2)} &\geq 2^{R(2-\epsilon)} \\ P' + N_1 &\geq \frac{N_1(P' + N_1 + N_2)}{N_1 + N_2}2^{R(2-\epsilon)} \\ P'\left[1 - \frac{N_1}{N_1 + N_2}2^{R(2-\epsilon)}\right] &\geq N_1 2^{R(2-\epsilon)} - N_1 \\ P' = \frac{P}{1 + 2\delta\ln(2)} &\geq N_1(N_1 + N_2)\left[\frac{2^{R(2-\epsilon)} - 1}{N_1 + N_2 - N_1 2^{R(2-\epsilon)}}\right] \\ 1 + 2\delta\ln(2) &\leq \frac{P}{N_1(N_1 + N_2)}\left[\frac{N_1 + N_2 - N_1 2^{R(2-\epsilon)}}{2^{R(2-\epsilon)} - 1}\right] \end{aligned}$$

for

$$\delta \leq \delta_1 = \frac{1}{2\ln(2)}\left\{\frac{P}{N_1(N_1 + N_2)}\left[\frac{N_1 + N_2 - N_1 2^{R(2-\epsilon)}}{2^{R(2-\epsilon)} - 1}\right] - 1\right\}.$$

For the above condition to be valid, δ_1 must be positive, which requires that

$$\begin{aligned} \frac{P}{N_1(N_1 + N_2)}\left[\frac{N_1 + N_2 - N_1 2^{R(2-\epsilon)}}{2^{R(2-\epsilon)} - 1}\right] &> 1 \\ P[N_2 + N_1 - N_1 2^{R(2-\epsilon)}] &> N_1(N_1 + N_2)[2^{R(2-\epsilon)} - 1] \\ PN_2 - PN_1[2^{R(2-\epsilon)} - 1] &> N_1(N_1 + N_2)[2^{R(2-\epsilon)} - 1] \\ PN_2 &> N_1(P + N_1 + N_2)[2^{R(2-\epsilon)} - 1]. \end{aligned}$$

The behavior of the right-hand side term in the above inequality depends on $R(2-\epsilon) = [I(\tilde{X};\tilde{Y}) - I(\tilde{X};\tilde{Z}) - \epsilon](2-\epsilon)$, which is parabolic in ϵ. When $\epsilon = 0$ or $2 + I(\tilde{X};\tilde{Y}) - I(\tilde{X};\tilde{Z})$,

$$\begin{aligned} R(2-\epsilon) &= [I(\tilde{X};\tilde{Y}) - I(\tilde{X};\tilde{Z}) - \epsilon](2-\epsilon) \\ &= 2[I(\tilde{X};\tilde{Y}) - I(\tilde{X};\tilde{Z})] \\ &= \log\left[\frac{(P + N_1)(N_1 + N_2)}{N_1(P + N_1 + N_2)}\right], \end{aligned}$$

which makes

$$\begin{aligned} N_1(P+N_1+N_2)[2^{R(2-\epsilon)}-1] &= N_1(P+N_1+N_2)\left[\frac{(P+N_1)(N_1+N_2)}{N_1(P+N_1+N_2)}-1\right] \\ &= (P+N_1)(N_1+N_2)-N_1(P+N_1+N_2) \\ &= PN_1+PN_2+N_1^2+N_1N_2-PN_1-N_1^2-N_1N_2 \\ &= PN_2. \end{aligned}$$

For $\epsilon \in (0, 2+I(\tilde{X};\tilde{Y})-I(\tilde{X};\tilde{Z}))$, $R(2-\epsilon) < 2[I(\tilde{X};\tilde{Y})-I(\tilde{X};\tilde{Z})]$, and $PN_2 > N_1(P+N_1+N_2)[2^{R(2-\epsilon)}-1]$ as required. Note that, for $\epsilon \geq 2+I(\tilde{X};\tilde{Y})-I(\tilde{X};\tilde{Z})$, $I(\tilde{X};\tilde{Y})-I(\tilde{X};\tilde{Z})-\epsilon < 0$, $1-\epsilon < 0$, $\epsilon > 1$, and there is nothing to prove.

In step 3, the wiretapper's decoding process is considered. In the wiretapper's decoding process, the wiretapper must select a parameter ξ for the joint typical decoder. If $\xi \neq \delta$, the set $T_X^N(\delta) \neq T_X^N(\xi)$, and the decoder is not reliable. Hence, the wiretapper should select $\xi = \delta$, and that is what we will assume in the following analysis.

Firstly, the entropy of the codeword conditioned on bin j and the wiretapper's observation $H(X_c|S^K, Z^N)$ is related to the probability of error P_B in the wiretapper's bin-decoding process through random variable χ defined as

$$\chi = \begin{cases} 1 & \text{if } \psi(Z^N) \neq X_c \\ 0 & \text{if } \psi(Z^N) = X_c \end{cases},$$

where $\psi(Z^N)$ is the wiretapper's decoding function. In the bin-decoding process, the wiretapper detects the codeword based on the knowledge of the bin, from which the codeword was selected, and the observation from the wiretap channel.

The probability of error in the wiretapper's bin-decoding process is then bounded using the union bound resulting in constraints on δ and N.

Now, the entropy of the codeword conditioned on bin j and the wiretapper's observation Z^N

is related to the probability of error in the bin-decoding process.

$$\begin{aligned}
H(X_c|S^K,Z^N) &= H(\chi,X_c|S^K,Z^N)-H(\chi|X_c,S^K,Z^N)\\
&= H(\chi|S^K,Z^N)+H(X_c|\chi,S^K,Z^N)-H(\chi|X_c,S^K,Z^N)\\
&\overset{(15)}{\leq} H(\chi|S^K)+H(X_c|\chi,S^K,Z^N)\\
&= \sum_{j=1}^{2^{NR}}\Pr\{S^K=j\}[H(\chi|S^K=j)\\
&\quad +p(\chi=0|S^K=j)H(X_c|Z^N,S^K=j,\chi=0)\\
&\quad +p(\chi=1|S^K=j)H(X_c|Z^N,S^K=j,\chi=1)]\\
&\overset{(16)}{=} \sum_{j=1}^{2^{NR}}\Pr\{S^K=j\}[h(P_B)+(1-P_B)H(X_c|Z^N,S^K=j,\chi=0)+\\
&\quad P_BH(X_c|Z^N,S^K=j,\chi=1)]\\
&\overset{(17)}{\leq} \sum_{j=1}^{2^{NR}}\Pr\{S^K=j\}[h(P_B)+P_B\log(2^{N[I(\tilde{X};\tilde{Z})-\epsilon]}-1)]\\
&\overset{(18)}{=} h(P_B)+P_B\log(2^{N[I(\tilde{X};\tilde{Z})-\epsilon]}-1)\\
&\leq h(P_B)+P_BN[I(\tilde{X};\tilde{Z})-\epsilon]\\
&\leq h(P_B)+P_BNI(\tilde{X};\tilde{Z})\\
\frac{H(X_c|S^K,Z^N)}{RN} &\leq \frac{h(P_B)+P_BNI(\tilde{X};\tilde{Z})}{RN}\\
&\leq \epsilon/2,
\end{aligned}$$

for $N \geq L_3 = 2h(P_B)/\{R\epsilon - 2P_BI(\tilde{X};\tilde{Z})\}$, and $P_B < R\epsilon/2I(\tilde{X};\tilde{Z})$, where $h(P_B)$ is the binary entropy function. (15) follows from the fact that $H(\chi|S^K,Z^N) \leq H(\chi|S^K)$ and $H(\chi,|X_c,S^K,Z^N) = 0$. (16) results from letting $P_B = p(\chi=1|S^K=j)$. (17) follows from the fact that $H(X_c|Z^N,S^K=j,\chi=0)=0$ and $H(X_c|Z^N,S^K=j,\chi=1)$ is less than log of the number of incorrect codewords in the bin. (18) follows from the fact that P_B is independent of the message.

To bound P_B, define the following events associated with the wiretapper's bin-decoding process:

- $\mathcal{E}^{Z1}(j)$: in the wiretapper's bin-decoding process, there is no sequence x_c in bin j that is jointly typical the received sequence z^N.
- $\mathcal{E}^{Z2}(j)$: in the wiretapper's bin-decoding process, a sequence $x_c(j,k) \neq x_c(j,k^*)$ is jointly typical with the received sequence z^N for $k \neq k^*$.

An error in wiretapper's bin-decoding process occurs when at least one of the events $\mathcal{E}^{Z1}(j)$ and $\mathcal{E}^{Z2}(j)$ occurs when there is no encoding error. Then, the probability of wiretapper's bin-decoding error given the secret message is

$$\begin{aligned}
P_B &= p(\chi=1|S^K=j)\\
&\leq \Pr\{\mathcal{E}^{Z1}(j)|\mathcal{E}^X(j)^C\}+\Pr\{\mathcal{E}^{Z2}(j)|\mathcal{E}^X(j)^C\}\\
&= \Pr\{\mathcal{E}^{Z1}(1)|\mathcal{E}^X(1)^C\}+\Pr\{\mathcal{E}^{Z2}(1)|\mathcal{E}^X(1)^C\}.
\end{aligned}$$

The two probabilities, $\Pr\{\mathcal{E}^{Z1}(1)|\mathcal{E}^X(1)^C\}$ and $\Pr\{\mathcal{E}^{Z2}(1)|\mathcal{E}^X(1)^C\}$, are evaluated as follows:

$\Pr\{\mathcal{E}^{Z1}(1)|\mathcal{E}^X(1)^C\}$. By Lemma 2.5 with $\epsilon_1 = R\epsilon/6I(\tilde{X};\tilde{Z})$, there exists an integer L_4 such that $\Pr\{(X_c, Z^N) \in T^N_{X,Z}(\delta)\} > 1 - R\epsilon/6I(\tilde{X};\tilde{Z})$ for given $\epsilon > 0$ and $\delta > 0$ and $N \geq L_4$, which implies that $\Pr\{\mathcal{E}^{Z1}(1)|\mathcal{E}^X(1)^C\} \leq R\epsilon/6I(\tilde{X};\tilde{Z})$ for given $\epsilon > 0$ and $\delta > 0$ and $N \geq L_4$.

$\Pr\{\mathcal{E}^{Z2}(1)|\mathcal{E}^X(1)^C\}$. By the code generating process $X_c(1,k^*)$ and $X_c(1,k)$ are independent, and so are Z^N and $X_c(1,k)$ for $k \neq k^*$. By inequality (2.20), $\Pr\{(X_c(1,k), Z^N) \in T^N_{XZ}(\delta) < 2^{-N[I(X;Z)-3\delta]}$ for $k \neq k^*$. Hence,

$$\begin{aligned}
\Pr\{\mathcal{E}^{Z2}(1)|\mathcal{E}^X(1)^C\} &\leq \sum_{k\neq k^*} 2^{-N[I(X;Z)-3\delta]} \\
&\leq (2^{N[I(\tilde{X};\tilde{Z})-\epsilon]} - 1)2^{-N[I(X;Z)-3\delta]} \\
&< 2^{-N[I(X;Z)-3\delta-I(\tilde{X};\tilde{Z})+\epsilon]} \\
&\leq R\epsilon/6I(\tilde{X};\tilde{Z}),
\end{aligned}$$

when

$$I(X;Z) - I(\tilde{X};\tilde{Z}) - 3\delta + \epsilon \;>\; 0,$$

and

$$\begin{aligned}
N[I(X;Z) - I(\tilde{X};\tilde{Z}) - 3\delta + \epsilon] &\geq \log\left[\frac{6I(\tilde{X};\tilde{Z})}{R\epsilon}\right] \\
N &\geq L_5,
\end{aligned}$$

where

$$L_5 = \frac{1}{I(X;Z) - I(\tilde{X};\tilde{Z}) - 3\delta + \epsilon} \log\left[\frac{6I(\tilde{X};\tilde{Z})}{R\epsilon}\right].$$

Now, let $g_{LHS}(\delta) = I(X;Z) - I(\tilde{X};\tilde{Z})$, and $g_{RHS}(\delta) = 3\delta - \epsilon$. We then have that

$$\begin{aligned}
g_{LHS}(\delta) &\overset{(19)}{=} \frac{1}{2}\log\left[\frac{P(1+2\delta\ln(2))^{-1} + N_1 + N_2}{N_1 + N_2}\right] - \frac{1}{2}\log\left[\frac{P + N_1 + N_2}{N_1 + N_2}\right] \\
&= \frac{1}{2}\log\left[\frac{P(1+2\delta\ln(2))^{-1} + N_1 + N_2}{P + N_1 + N_2}\right] \\
g_{LHS}(0) &= 0 \\
\lim_{\delta\to\infty} g_{LHS}(\delta) &= \frac{1}{2}\log\left[\frac{N_1 + N_2}{P + N_1 + N_2}\right] \\
\frac{dg_{LHS}(\delta)}{d\delta} &= \frac{-P(1+2\delta)^{-1}}{P + (N_1 + N_2)(1+2\delta)} \\
&< 0 \text{ because } P, N_1, N_2, \delta > 0.
\end{aligned}$$

(19) follows from equation (2.13). From the facts that $g_{LHS}(0) = 0$, $g_{RHS}(0) = -\epsilon < 0$, and $g_{LHS}(\delta)$ is a monotonically decreasing continuous function converging to a negative number

while $g_{RHS}(\delta)$ is a monotonically increasing continuous function approaching ∞ as δ increases, it follows that there exists a $\delta_2 > 0$ such that $g_{LHS}(\delta_2) = g_{RHS}(\delta_2)$ by the continuity of the functions [13]. Thus, $g_{LHS}(\delta) - g_{RHS}(\delta) = I(X;Z) - I(\tilde{X};\tilde{Z}) - 3\delta + \epsilon > 0$ for $0 < \delta < \delta_2$, and

$$\begin{aligned} P_B &\leq \Pr\{\mathcal{E}^{Z1}(1)|\mathcal{E}^X(1)^C\} + \Pr\{\mathcal{E}^{Z2}(1)|\mathcal{E}^X(1)^C\} \\ &\leq R\epsilon/6 I(\tilde{X};\tilde{Z}) + R\epsilon/6 I(\tilde{X};\tilde{Z}) \\ &= R\epsilon/3 I(\tilde{X};\tilde{Z}) \\ &< R\epsilon/2 I(\tilde{X};\tilde{Z}), \end{aligned}$$

for $0 < \delta < \delta_2$, and $N \geq \max\{L_4, L_5\}$, and $H(X_c|S^K, Z^N)/(RN) \leq \epsilon/2$ when $N \geq L_3$; hence, the rate-equivocation pair $(C_s, 1)$ is achievable for $0 < \delta < \min\{\delta_0, \delta_1, \delta_2\}$, and $N \geq N_{min} = \max\{L_0, L_1, L_2, L_3, L_4, L_5\}$. ■

By Lemma 4.13, time-sharing the code in the interference-free mode with the a capacity achieving code (without the wiretapper) for the main channel achieves all the rate-equivocation pairs in the region $R \leq C_M, d \leq 1$, and $Rd \leq C_s$, where C_M is the capacity of the main channel, and $C_s = C_M - C_{MW}$, and C_{MW} is the capacity of the combined channel as shown in [14].

Lemma 4.13 *[14] Let $R_1 d_1 = R_2 d_2 = c$, a constant. Assume $R_1 > R_2$ and hence $d_1 < d_2$. If the points (R_1, d_1) and (R_2, d_2) are achievable, then by time-sharing any point (R, d) with $R_2 \leq R \leq R_1, d_1 \leq d \leq d_2$, and $Rd = c$ is achievable.*

4.3 Gaussian Wiretap Channel with Side Information

The Gaussian wiretap channel model is extended to the Gaussian wiretap channel with side information. We propose the camouflage and the high-power modes by extending the interference-free mode to take into account the effect of the interference in the channel. The rate-equivocation pairs achieved by these two strategies and the result of time-sharing codes in Lemma 4.13 lead to an achievable region for GWCSI.

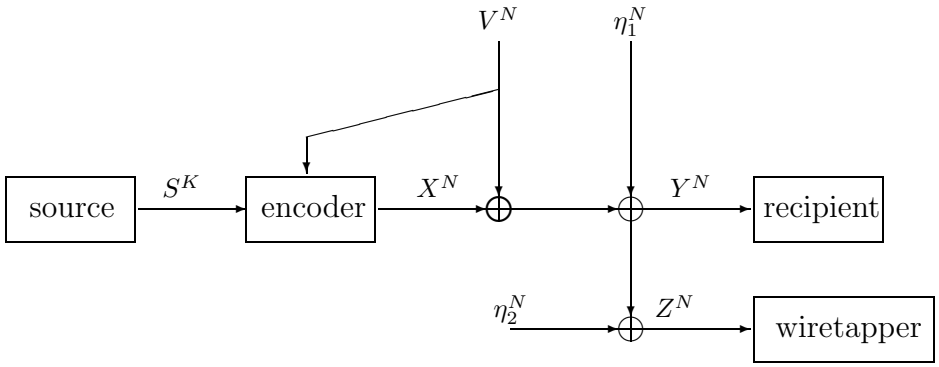

Figure 4.3: The Gaussian wiretap channel with side information.

4.3.1 Model Description

GWCSI is extended from GWC by adding an interference in the main channel. The interference is modelled as a sequence V^N of i.i.d. random variables such that $V \sim \mathcal{N}(0, Q)$, independent of the noises and the codeword. We assume that the whole sequence of the interference is known to the encoder before the secret message transmission. We note that the main channel of GWCSI is the same as the dirty-paper channel investigated in [3]. The modes of operation will be given based on the leakage function, defined in the following section, for GWCSI.

4.3.2 Leakage Function

Gel'fand and Pinsker proved the capacity region for the channel with random parameters in [12] by using an auxiliary random variable U to convey information about the message and the interference to the recipient. The average information rate between the encoder and the recipient is $I(U;Y) - I(U;V)$, which can be interpreted as the average rate of information about the message and the interference at the recipient less the average rate of information about the interference contained in the auxiliary random variable U.

The dirty-paper channel investigated by Costa in [3] is a special case of the channel with random parameters in that the interference and the noise are modelled as independent sequences of i.i.d. Gaussian random variables. In the coding strategy proposed by Costa, an auxiliary random variable $U = X + \alpha V$ is also used to convey the information about both the secret message and the interference to the recipient, where α is a real number.

In the GWCSI model, the main channel is the same as the dirty-paper channel while the combined channel is a physically degraded version of the main channel. If we use an auxiliary random variable U to convey the secret message and the interference to the recipient, then $I(U;Y) - I(U;V)$ is the average information rate between the encoder and the recipient, and $I(U;Z) - I(U;V)$ is that between the encoder and the wiretapper. Consequently, $I(U;Z) - I(U;V)$ can be considered as the average information leakage rate to the wiretapper. Note that when there is no interference in the main channel, and the auxiliary random variable $U = X$, $I(U;Z) - I(U;V)$ reduces to $I(X;Z)$ as defined in the GWC case.

Calculating the Leakage Function for GWCSI

Similar to the coding strategy for the dirty-paper channel, the camouflage and high-power modes make use of an auxiliary random variable $U = X + \alpha V$ to carry information about the secret message and the interference. The two modes of operation are defined by codebook generating, encoding and decoding processes. In the description of the two modes, we will need the following random variables:

$\tilde{U} \sim \mathcal{N}(0, P + \alpha^2 Q)$, $\tilde{Y} \sim \mathcal{N}(0, P + Q + N_1)$, $\tilde{Z} \sim \mathcal{N}(0, P + Q + N_1 + N_2)$
$U \sim \mathcal{N}(0, P' + \alpha^2 Q)$, $Y \sim \mathcal{N}(0, P' + Q + N_1)$, $Z \sim \mathcal{N}(0, P' + Q + N_1 + N_2)$
$U = X + \alpha V$
U_c represents the output of the auxiliary codeword selection process,
where $P' = P(1 + 4\delta \ln(2))^{-1} < P$ and $\delta > 0$.

Let $\tilde{U}^N$, $\tilde{Y}^N$, $\tilde{Z}^N$ be sequences of i.i.d. Gaussian random variables $\tilde{U}$, $\tilde{Y}$ and $\tilde{Z}$, respectively. We can evaluate the leakage function $\Delta I(\alpha) = I(\tilde{U};\tilde{Z}) - I(\tilde{U};V)$ by calculating the correlation coefficients $\varrho_{\tilde{U}\tilde{Z}}$ and $\varrho_{\tilde{U}V}$, and $I(\tilde{U};\tilde{Z})$ and $I(\tilde{U};V)$, respectively. By the definition of the correlation coefficient,

$$\begin{aligned}
\varrho_{\tilde{U}\tilde{Z}} &= \frac{E[\tilde{U}\tilde{Z}]}{\sigma_{\tilde{U}}\sigma_{\tilde{Z}}} \\
&= \frac{E[(\tilde{X}+\alpha V)(\tilde{X}+V+\eta_1+\eta_2)]}{\sigma_{\tilde{U}}\sigma_{\tilde{Z}}} \\
&= \frac{P+\alpha Q}{\sigma_{\tilde{U}}\sigma_{\tilde{Z}}} \\
\varrho^2_{\tilde{U}\tilde{Z}} &= \frac{(P+\alpha Q)^2}{(P+\alpha^2 Q)(P+Q+N_1+N_2)}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\varrho^2_{\tilde{U}V} &= \frac{E[(\tilde{X}+\alpha V)V]^2}{\sigma^2_{\tilde{U}}\sigma^2_V} \\
&= \frac{\alpha^2\sigma^2_V}{\sigma^2_{\tilde{U}}} = \frac{\alpha^2 Q}{P+\alpha^2 Q}.
\end{aligned}$$

By equation (2.11), we have

$$\begin{aligned}
I(\tilde{U};\tilde{Z}) &= \frac{1}{2}\log\left[\frac{1}{1-\varrho^2_{\tilde{U}\tilde{Z}}}\right] \\
&= \frac{1}{2}\log\left[\frac{(P+\alpha^2Q)(P+Q+N_1+N_2)}{(P+\alpha^2Q)(P+Q+N_1+N_2)-(P+\alpha Q)^2}\right] \qquad (4.1) \\
I(\tilde{U};V) &= \frac{1}{2}\log\left[\frac{P+\alpha^2Q}{(P+\alpha^2Q)-\alpha^2Q}\right] \\
&= \frac{1}{2}\log\left[\frac{P+\alpha^2Q}{P}\right] \qquad (4.2)
\end{aligned}$$

The leakage function is then

$$\begin{aligned}
\Delta I(\alpha) &= \frac{1}{2}\log\left[\frac{(P+\alpha^2Q)(P+Q+N_1+N_2)}{(P+\alpha^2Q)(P+Q+N_1+N_2)-(P+\alpha Q)^2}\right] - \frac{1}{2}\log\left[\frac{P+\alpha^2Q}{P}\right] \\
&= \frac{1}{2}\log\left[\frac{P(P+Q+N_1+N_2)}{(P+\alpha^2Q)(P+Q+N_1+N_2)-(P+\alpha Q)^2}\right]. \qquad (4.3)
\end{aligned}$$

If the leakage function is positive, some information communicated between the encoder and the recipient is leaked to the wiretapper; otherwise, the information is kept secret from the wiretapper. Correspondingly, the camouflage mode is designed for the case when the leakage is non-positive while the high-power mode for the case when the leakage is positive.

4.3.3 Camouflage Mode

We now define the camouflage mode and prove that it can be used to achieve the rate-equivocation pair $(C_M, 1)$ if $\Delta I(\alpha) \leq 0$. Given an $\epsilon > 0$, let ϵ_{UV} and ϵ_{UZ} be constants associated with the codebook generating process, α and δ be constants associated with the encoding and decoding processes, and ξ be a constant associated with the wiretapper's decoding process to be specified later. For $\Delta I(\alpha) \leq 0$,

1. *Generating the codebook.* Generate $2^{N[I(\tilde{U};\tilde{Y})-\epsilon+\epsilon_{UV}]}$ sequences u^N according to the distribution $p(u^N) = \prod_{i=1}^{N} p(u_i)$, and $p(u_i) \sim \mathcal{N}(0, P' + \alpha^2 Q)$ for all $i \in \{1, 2, \ldots, N\}$. Place the sequences u^N randomly into $2^{N[I(\tilde{U};\tilde{Y})-I(\tilde{U};V)-\epsilon]}$ bins, each of which contains $2^{N[I(\tilde{U};V)-I(\tilde{U};\tilde{Z})+\epsilon_{UV}+\epsilon_{UZ}]}$ subbins so that each subbin contains $2^{N[I(\tilde{U};\tilde{Z})-\epsilon_{UZ}]}$ sequences, and each bin contains $2^{N[I(\tilde{U};V)+\epsilon_{UV}]}$ sequences. Index each bin by $j \in \{1, 2, \ldots, 2^{NR}\}$, and each subbin by $w \in \{1, 2, \ldots, 2^{N[I(\tilde{U};V)-I(\tilde{U};\tilde{Z})+\epsilon_{UV}+\epsilon_{UZ}]}\}$, where $R = I(\tilde{U};\tilde{Y}) - I(\tilde{U};V) - \epsilon$. The codebook is given to the sender and the recipient.

2. *Encoding.* To send a message j through an interference v^N, the sender looks for a sequence u_c in bin j such that $(u_c, v^N) \in T^N_{U,V}(\delta)$ and transmit $x^N = u_c - \alpha v^N$. If there is more than one sequence u_c that is jointly typical with v^N, randomly select one.

3. *Decoding.* To decode the received sequence for the message, the recipient finds a sequence u_c in the codebook that is jointly typical with the received sequence y^N, i.e., $(u_c, y^N) \in T^N_{U,Y}(\delta)$. Declare the index to the bin, in which the sequence is found as the received message.

4. *Wiretapper's decoding.* The wiretapper receives a sequence z^N and finds a sequence u_c in the codebook that is jointly typical with the received sequence, i.e., $(u_c, z^N) \in T^N_{U,Z}(\xi)$. Declare the index to the bin, in which the sequence is found as the received message.

5. *Probability of error.* An error occurs in the coding process when a message j is to be transmitted and one or more of the following events occurs.

   - $\mathcal{E}^V(j)$: in the encoding process, there is no sequence u_c in bin j that is jointly typical with the interference sequence.
   - $\mathcal{E}^X(j)$: in the encoding process, $x^N = u_c - \alpha v^N$ does not satisfy the power constraint.
   - $\mathcal{E}^{Y1}(j)$: in the decoding process, there is no sequence u_c that is jointly typical with the received sequence.
   - $\mathcal{E}^{Y2}(j)$: in the decoding process, a sequence u_c in bin $i \neq j$ is jointly typical with the received sequence.

Theorem 4.2 *Given the camouflage mode of operation to be used in the Gaussian wiretap channel with side information and average power constraint P, for any $\epsilon > 0$ and*

$$\begin{aligned} N_2 &> N_1 + \frac{N_1^2}{Q}, \\ 0 \leq P &\leq -N_1 - \frac{Q}{2} + \frac{\sqrt{Q^2 + 4QN_2}}{2}, \end{aligned}$$

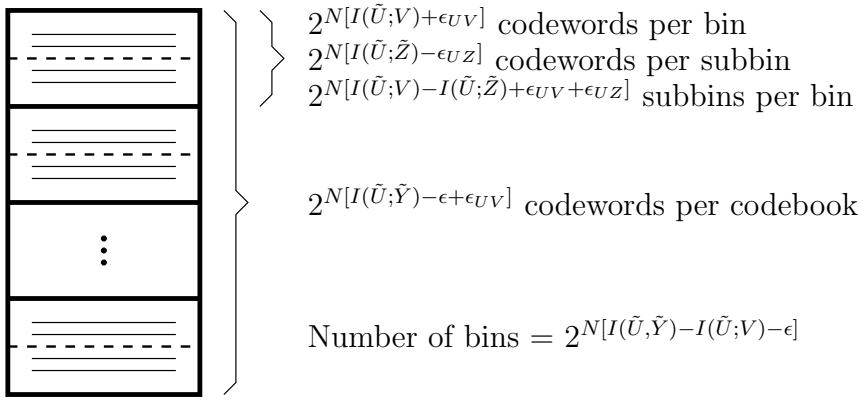

Figure 4.4: The codebook for the camouflage mode.

there exists an integer N_{min} such that, for $N \geq N_{min}$,

$$\begin{aligned} \frac{H(S^K)}{N} &\geq R_{cam} \triangleq \frac{1}{2}\log\left[\frac{P+N_1}{N_1}\right] - \epsilon \\ \frac{H(S^K|Z^N)}{H(S^K)} &\geq 1-\epsilon \\ P_e &\leq \epsilon. \end{aligned}$$

Proof: This proof consists of three main parts: proof of the rate, proof of the probability of error, and proof of the equivocation.

Proof of the Rate. The number of messages that can be sent is the number of bins in the codebook, $2^{NR_{cam}(\alpha)} = 2^{N[I(\tilde{U};\tilde{Y})-I(\tilde{U};V)-\epsilon]}$. We evaluate $\varrho_{\tilde{U}\tilde{Y}}$, $I(\tilde{U};\tilde{Y})$ and $R_{cam}(\alpha) = [I(\tilde{U};\tilde{Y}) - I(\tilde{U};V) - \epsilon]$ in that order. The correlation coefficient is

$$\begin{aligned} \varrho_{\tilde{U}\tilde{Y}} &= \frac{E[\tilde{U}\tilde{Y}]}{\sigma_{\tilde{U}}\sigma_{\tilde{Y}}} \\ &= \frac{E[(\tilde{X}+\alpha V)(\tilde{X}+V+\eta_1)]}{\sigma_{\tilde{U}}\sigma_{\tilde{Y}}} \\ \varrho^2_{\tilde{U}\tilde{Y}} &= \frac{(P+\alpha Q)^2}{(P+\alpha^2 Q)(P+Q+N_1)} \\ \frac{1}{1-\varrho^2_{\tilde{U}\tilde{Y}}} &= \frac{(P+\alpha^2 Q)(P+Q+N_1)}{(P+\alpha^2 Q)(P+Q+N_1)-(P+\alpha Q)^2}, \end{aligned}$$

and

$$I(\tilde{U};\tilde{Y}) \stackrel{(1)}{=} \frac{1}{2}\log\left[\frac{(P+\alpha^2Q)(P+Q+N_1)}{(P+\alpha^2Q)(P+Q+N_1)-(P+\alpha Q)^2}\right] \tag{4.4}$$

$$\begin{aligned} I(\tilde{U};\tilde{Y}) - I(\tilde{U};V) &\stackrel{(2)}{=} \frac{1}{2}\log\left[\frac{(P+\alpha^2Q)(P+Q+N_1)}{(P+\alpha^2Q)(P+Q+N_1)-(P+\alpha Q)^2}\right] \\ &\quad -\frac{1}{2}\log\left[\frac{P+\alpha^2Q}{P}\right] \\ &= \frac{1}{2}\log\left[\frac{P(P+Q+N_1)}{(P+\alpha^2Q)(P+Q+N_1)-(P+\alpha Q)^2}\right]. \end{aligned} \tag{4.5}$$

(1) follows from equation (2.11). (2) follows from equation (4.2).

Hence,

$$\begin{aligned} R_{cam}(\alpha) &= I(\tilde{U};\tilde{Y}) - I(\tilde{U};V) - \epsilon \\ &\stackrel{(3)}{=} \frac{1}{2}\log\left[\frac{P(P+Q+N_1)}{(P+\alpha^2Q)(P+Q+N_1)-(P+\alpha Q)^2}\right] - \epsilon. \end{aligned} \tag{4.6}$$

(3) follows from equation (4.5). We select the optimal α for this coding strategy by considering

$$\begin{aligned} \frac{dR_{cam}(\alpha)}{d\alpha} &= \frac{1}{2\ln(2)}\left[\frac{(P+\alpha^2Q)(P+Q+N_1)-(P+\alpha Q)^2}{P(P+Q+N_1)}\right]\times \\ &\quad -\left[\frac{\{P(P+Q+N_1)\}\{2\alpha Q(P+Q+N_1)-2Q(P+\alpha Q)\}}{\{(P+\alpha^2Q)(P+Q+N_1)-(P+\alpha Q)^2\}^2}\right] \\ &= -\frac{1}{2\ln(2)}\left[\frac{2\alpha Q(P+Q+N_1)-2Q(P+\alpha Q)}{(P+\alpha^2Q)(P+Q+N_1)-(P+\alpha Q)^2}\right]. \end{aligned}$$

Setting $dR_{cam}(\alpha)/d\alpha = 0$, we have

$$\begin{aligned} 2\alpha Q(P+Q+N_1) &= 2Q(P+\alpha Q) \\ \alpha(P+N_1)+\alpha Q &= P+\alpha Q \\ \alpha &= \alpha_{cam} = \frac{P}{P+N_1}, \end{aligned} \tag{4.7}$$

and

$$\begin{aligned}
R_{cam}(\alpha_{cam}) &\overset{(4)}{=} \frac{1}{2}\log\left[\frac{P(P+Q+N_1)}{(P+\alpha_{cam}^2 Q)(P+Q+N_1)-(P+\alpha_{cam}Q)^2}\right]-\epsilon \\
&= \frac{1}{2}\log\left[\frac{P(P+N_1)^2(P+Q+N_1)}{(P(P+N_1)^2+P^2Q)(P+Q+N_1)-(P(P+N_1)+PQ)^2}\right]-\epsilon \\
&= \frac{1}{2}\log\left[\frac{(P+N_1)^2}{(P+N_1)^2+PQ-P(P+Q+N_1)}\right]-\epsilon \\
&= \frac{1}{2}\log\left[\frac{P+N_1}{(P+N_1)-P)}\right]-\epsilon \\
&= \frac{1}{2}\log\left[\frac{P+N_1}{N_1}\right]-\epsilon. \qquad (4.8)
\end{aligned}$$

(4) follows from equation (4.6).

Proof of the probability of error. The probability of error is bounded by the probabilities of $\{\mathcal{E}^V(1)\}$ $\{\mathcal{E}^X(1)|\mathcal{E}^V(1)^C\}$, $\{\mathcal{E}^{Y1}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}$, and $\{\mathcal{E}^{Y2}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}$, each of which can be bounded in the following steps.

Denote the k^{th} auxiliary codeword in bin j by $u_c(j,k)$ and the selected auxiliary codeword in bin j, which is jointly typical with the interference, by $u_c(j,k^*)$, where $k \in \{1,2,\ldots,2^{N[I(\tilde{U};V)+\epsilon_{UV}]}\}$. The transmitted codeword is then $x^N = u_c(j,k^*) - \alpha v^N$.

$$\begin{aligned}
P_e &\overset{(5)}{\leq} \sum_{j=1}^{2^{NR_{cam}(\alpha)}} \Pr\{S^K=j\}[\Pr\{\mathcal{E}^V(j)\}+\Pr\{\mathcal{E}^X(j)|\mathcal{E}^V(j)^C\} \\
&\quad +\Pr\{\mathcal{E}^{Y1}(j)|\mathcal{E}^X(j)^C\mathcal{E}^V(j)^C\}+\Pr\{\mathcal{E}^{Y2}(j)|\mathcal{E}^X(j)^C\mathcal{E}^V(j)^C\}] \\
&\overset{(6)}{=} \Pr\{\mathcal{E}^V(1)\}+\Pr\{\mathcal{E}^X(1)|\mathcal{E}^V(1)^C\} \\
&\quad +\Pr\{\mathcal{E}^{Y1}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}+\Pr\{\mathcal{E}^{Y2}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}].
\end{aligned}$$

(5) follows from the union bound. (6) follows from the fact that error events don't depend on the message j. We now consider the four terms in the summation separately.

$\Pr\{\mathcal{E}^V(1)\}$. By the code generating process, U_c and V^N are independent. By equation (2.21),

$$\begin{aligned}
\Pr\{(U_c,V^N)\notin T^N_{U,V}(\delta)\} &\leq 1-(1-\delta)2^{-N[I(U;V)+3\delta]} \\
\Pr\{\mathcal{E}^V(1)\} &\overset{(7)}{\leq} [1-(1-\delta)2^{-N[I(U;V)+3\delta]}]^{2^{N[I(\tilde{U};V)+\epsilon_{UV}]}} \\
&\overset{(8)}{\leq} \exp\{-(1-\delta)2^{-N[I(U;V)+3\delta]}\}^{2^{N[I(\tilde{U};V)+\epsilon_{UV}]}} \\
&= \exp\{-(1-\delta)2^{N[I(\tilde{U};V)-I(U;V)+\epsilon_{UV}-3\delta]}\} \\
&\leq \epsilon/3
\end{aligned}$$

when $\delta < 1$, and

$$I(\tilde{U};V)-I(U;V)+\epsilon_{UV}-3\delta > 0 \qquad (4.9)$$

and

$$\begin{aligned}(1-\delta)2^{N[I(\tilde{U};V)-I(U;V)+\epsilon_{UV}-3\delta]} &\geq \ln\left[\frac{3}{\epsilon}\right] \\ N[I(\tilde{U};V)-I(U;V)+\epsilon_{UV}-3\delta] &\geq \log\left[\frac{1}{1-\delta}\ln\left[\frac{3}{\epsilon}\right]\right] \\ N &\geq L_0,\end{aligned}$$

where

$$L_0 = \frac{1}{I(\tilde{U};V)-I(U;V)+\epsilon_{UV}-3\delta}\log\left[\frac{1}{1-\delta}\ln\left[\frac{3}{\epsilon}\right]\right].$$

(7) follows from the fact that there are $2^{N[I(\tilde{U};V)+\epsilon_{UV}]}$ codewords in a bin. (8) follows from the fact that $e^a \geq 1+a$ (Taylor's expansion of e^a) [18].

We now consider the condition when inequality (4.9) is satisfied as follows.

$$\begin{aligned}I(U;V) &= \frac{1}{2}\log\left[\frac{P'+\alpha^2Q}{P'}\right] \\ I(\tilde{U};V)-I(U;V) &= \frac{1}{2}\log\left[\frac{P+\alpha^2Q}{P}\right]-\frac{1}{2}\log\left[\frac{P'+\alpha^2Q}{P'}\right] \\ &= \frac{1}{2}\log\left[\frac{(P+\alpha^2Q)P'}{P(P'+\alpha^2Q)}\right] \\ &= \frac{1}{2}\log\left[\frac{(P+\alpha^2Q)}{P+\alpha^2Q(1+4\delta\ln(2))}\right].\end{aligned}$$

Let $g_{LHS}(\delta) = I(\tilde{U};V)-I(U;V)$, and $g_{RHS}(\delta) = 3\delta-\epsilon_{UV}$. Then, $g_{LHS}(\delta)$ and $g_{RHS}(\delta)$ are continuous functions since the equations

$$\begin{aligned}P+\alpha^2Q &= 0 \\ P+\alpha^2Q(1+4\delta\ln(2)) &= 0\end{aligned}$$

have no real solution for α when $\delta, P, Q > 0$.

$$\begin{aligned}g_{LHS}(0) &= 0 \\ \lim_{\delta\to\infty} g_{LHS}(\delta) &= -\infty \\ \frac{dg_{LHS}(\delta)}{d\delta} &= \left[\frac{1}{2\ln(2)}\right]\left[\frac{P+\alpha^2Q(1+4\delta\ln(2))}{P+\alpha^2Q}\right]\left[\frac{-(P+\alpha^2Q)(4\ln(2)\alpha^2Q)}{\{P+\alpha^2Q(1+4\delta\ln(2))\}^2}\right] \\ &= \frac{-2\alpha^2Q}{P+\alpha^2Q(1+4\delta\ln(2))} \\ &< 0 \text{ for } \delta, P, Q > 0.\end{aligned}$$

From the facts that $g_{LHS}(0) = 0$, $g_{RHS}(0) = -\epsilon_{UV} < 0$, and $g_{LHS}(\delta)$ is a monotonically decreasing continuous function approaching $-\infty$ while $g_{RHS}(\delta)$ is a monotonically increasing

continuous function approaching ∞ as δ increases, it follows that there exists a $\delta_0 > 0$ such that $g_{LHS}(\delta_0) = g_{RHS}(\delta_0)$ by the continuity of the functions [13]. Thus, $g_{LHS}(\delta) - g_{RHS}(\delta) = I(\tilde{U};V) - I(U;V) + \epsilon_{UV} - 3\delta > 0$ for $0 < \delta < \delta_0$, and $\Pr\{\mathcal{E}^V(1)\} \leq \epsilon/3$ when $0 < \delta < \min\{1, \delta_0\}$ and $N \geq L_0$.

$\Pr\{\mathcal{E}^X(1)|\mathcal{E}^V(1)^C\}$. If the event $\mathcal{E}^V(1)$ does not occur, then we have that $(u_c(1,k^*), v^N) \in T^N_{U,V}(\delta)$. By Lemma 2.12, $\frac{1}{N}\sum_{i=1}^N x_i^2 \leq P$, and $\Pr\{\mathcal{E}^X(1)|\mathcal{E}^V(1)^C\} = 0$.

$\Pr\{\mathcal{E}^{Y1}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}$. By Lemma 2.5 with $\epsilon_1 = \epsilon/3$, there exists L_1 for a given $\delta > 0$ such that $\Pr\{(U_c, Y^N) \in T^N_{U,Y}(\delta)\} > 1 - \epsilon/3$ when $N \geq L_1$, which implies that $\Pr\{\mathcal{E}^{Y1}(1)|\mathcal{E}^X(1)^C \mathcal{E}^V(1)^C\} \leq \epsilon/3$ a given $\delta > 0$ and $N \geq L_1$.

$\Pr\{\mathcal{E}^{Y2}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}$. By the code generating process, $U_c(1,k^*)$ and $U_c(i,k)$ are independent, and so are Y^N and $U_c(i,k)$ for $(i,k) \neq (1,k^*)$. By inequality (2.20), $\Pr\{(U_c(i,k), Y^N) \in T^N_{U,Y}(\delta) < 2^{-N[I(U;Y)-3\delta]}$ for $(i,k) \neq (1,k^*)$. Hence,

$$\begin{aligned}
\Pr\{\mathcal{E}^{Y2}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}\} &< \sum_{(i,k)\neq(1,k^*)} 2^{-N[I(U;Y)-3\delta]} \\
&\overset{(9)}{\leq} (2^{N[I(\tilde{U};\tilde{Y})-\epsilon+\epsilon_{UV}]} - 1)2^{-N[I(U;Y)-3\delta]} \\
&< 2^{-N[I(U;Y)-3\delta-I(\tilde{U};\tilde{Y})+\epsilon-\epsilon_{UV}]} \\
&\leq \epsilon/3,
\end{aligned}$$

when

$$I(U;Y) - I(\tilde{U};\tilde{Y}) - 3\delta + \epsilon - \epsilon_{UV} > 0,$$

and

$$\begin{aligned}
N[I(U;Y) - I(\tilde{U};\tilde{Y}) - 3\delta + \epsilon - \epsilon_{UV}] &\geq \log\left[\frac{3}{\epsilon}\right] \\
N &\geq L_2,
\end{aligned}$$

where

$$L_2 = \frac{1}{I(U;Y) - I(\tilde{U};\tilde{Y}) - 3\delta + \epsilon - \epsilon_{UV}} \log\left[\frac{3}{\epsilon}\right].$$

(9) follows from the union bound.

Now, define $\lambda_Y(\delta) = I(U;Y) - I(\tilde{U};\tilde{Y}) - 3\delta + \epsilon - \epsilon_{UV}$, and $\phi(\delta) = (1 + 4\delta\ln(2))^{-1}$. We then have that

$$\begin{aligned}
\lambda_Y(\delta) &\overset{(10)}{=} \frac{1}{2}\log\left[\frac{(P\phi(\delta) + \alpha^2 Q)(P\phi(\delta) + Q + N_1)}{(P\phi(\delta) + \alpha^2 Q)(P\phi(\delta) + Q + N_1) - (P\phi(\delta) + \alpha Q)^2}\right] \\
&\quad - \frac{1}{2}\log\left[\frac{(P + \alpha^2 Q)(P + Q + N_1)}{(P + \alpha^2 Q)(P + Q + N_1) - (P + \alpha Q)^2}\right] - 3\delta + \epsilon - \epsilon_{UV} \\
\lambda_Y(0) &\overset{(11)}{>} 0 \text{ for } \epsilon > \epsilon_{UV} \qquad (4.10) \\
\lim_{\delta\to\infty} \lambda_Y(\delta) &= -\infty \text{ for finite } \alpha, P, Q, N_1 \qquad (4.11)
\end{aligned}$$

(10) follows from equation (4.4). (11) follows from the fact that $\phi(0) = 1$.

Furthermore, $\lambda_Y(\delta)$ is a continuous function because the equations

$$\begin{aligned}(P\phi(\delta) + \alpha^2 Q)(P\phi(\delta) + Q + N_1) &= 0\\ (P\phi(\delta) + \alpha^2 Q)(P\phi(\delta) + Q + N_1) - (P\phi(\delta) + \alpha Q)^2 &= 0\\ (P + \alpha^2 Q)(P + Q + N_1) &= 0\\ (P + \alpha^2 Q)(P + Q + N_1) - (P + \alpha Q)^2 &= 0\end{aligned}$$

have no real solutions for α when $P, Q, N_1, \delta > 0$.

Consequently, there exists the smallest $\delta_1 > 0$ such that $\lambda_Y(\delta_1) = 0$ by the continuity of $\lambda_Y(\delta)$ [13]. Thus, $\lambda_Y(\delta) = I(U;Y) - I(\tilde{U};\tilde{Y}) - 3\delta + \epsilon - \epsilon_{UV} > 0$ for $0 < \delta < \delta_1$ and $\epsilon > \epsilon_{UV}$, and $\Pr\{\mathcal{E}^{Y2}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\} \leq \epsilon/3$ for $0 < \delta < \delta_1$, $\epsilon > \epsilon_{UV}$ and $N \geq L_2$.

Combining the four terms of the summation,

$$P_e \leq \epsilon/3 + 0 + \epsilon/3 + \epsilon/3 = \epsilon,$$

for $0 < \delta < \min\{1, \delta_0, \delta_1\}$, $0 < \epsilon_{UV} < \epsilon$, and $N \geq \max\{L_0, L_1, L_2\}$.

Proof of the equivocation. The equivocation of the secret message is expressed in term of the probability of error in the wiretapper's subbin decoding. In doing so, let W ($w \in \{1, 2, \ldots, 2^{N[I(\tilde{U};V) - I(\tilde{U};\tilde{Z}) + \epsilon_{UV} + \epsilon_{UZ}]}\}$) be the random variable representing the outcome of the subbin selecting process and consider the uncertainty of the secret message to the wiretapper in three steps:

1. show that $H(S^K|Z^N) = N[I(U;Y) - I(U;Z) - I(\tilde{U};V) + I(\tilde{U};\tilde{Z}) - \epsilon_{UV} - \epsilon_{UZ}]$ $- H(U_c|S^K, W, Z^N)$;
2. show that $I(U;Y) - I(U;Z) - I(\tilde{U};V) + I(\tilde{U};\tilde{Z}) - \epsilon_{UV} - \epsilon_{UZ} \geq R_{cam}(\alpha)(1 - \epsilon/2)$;
3. show that $H(U_c|S^K, W, Z^N)/R_{cam}(\alpha)N \leq \epsilon/2$.

Combining the above three steps,

$$\begin{aligned}&\frac{H(S^K|Z^N)}{H(S^K)}\\ &\quad= \frac{N[I(U;Y) - I(U;Z) - I(\tilde{U};V) + I(\tilde{U};\tilde{Z}) - \epsilon_{UV} - \epsilon_{UZ}]}{R_{cam}(\alpha)N}\\ &\qquad - \frac{H(U_c|S^K, W, Z^N)}{R_{cam}(\alpha)N}\\ &\quad\geq 1 - \epsilon/2 - \frac{H(U_c|S^K, W, Z^N)}{R_{cam}(\alpha)N}\\ &\quad\geq 1 - \epsilon/2 - \epsilon/2\\ &\quad= 1 - \epsilon.\end{aligned}$$

We now proceed to step 1 by considering

$$\begin{aligned}
H(S^K|Z^N) &= H(S^K,Z^N)-H(Z^N)\\
&= H(S^K,W,Z^N)-H(W|S^K,Z^N)-H(Z^N)\\
&= H(S^K,W,U_c,Z^N)-H(U_c|S^K,W,Z^N)-H(W|S^K,Z^N)-H(Z^N)\\
&= H(S^K,W|U_c,Z^N)+H(U_c|Z^N)-H(U_c|S^K,W,Z^N)-H(W|S^K,Z^N)\\
&\overset{(12)}{=} H(U_c|Z^N)-H(U_c|S^K,W,Z^N)-H(W|S^K,Z^N)\\
&\overset{(13)}{\geq} H(U_c|Z^N)-H(U_c|Y^N)-H(W|S^K,Z^N)-H(U_c|S^K,W,Z^N)\\
&= H(U_c,Z^N)-H(Z^N)-H(U_c,Y^N)+H(Y^N)-H(W|S^K,Z^N)\\
&\quad -H(U_c|S^K,W,Z^N)\\
&= [H(U_c)+H(Z^N|U_c)]-H(Z^N)-[H(U_c)+H(Y^N|U_c)]+H(Y^N)\\
&\quad -H(W|S^K,Z^N)-H(U_c|S^K,W,Z^N)\\
&= H(Z^N|U_c)-[H(U^N,Z^N)-H(U^N|Z^N)]-H(Y^N|U_c)+[H(U^N,Y^N)\\
&\quad -H(U^N|Y^N)]-H(W|S^K,Z^N)-H(U_c|S^K,W,Z^N)\\
&= H(Z^N|U_c)-[H(U^N)+H(Z^N|U^N)-H(U^N|Z^N)]-H(Y^N|U_c)\\
&\quad +[H(U^N)+H(Y^N|U^N)-H(U^N|Y^N)]-H(W|S^K,Z^N)\\
&\quad -H(U_c|S^K,W,Z^N)\\
&\overset{(14)}{=} H(U^N|Z^N)-H(U^N|Y^N)-H(W|S^K,Z^N)-H(U_c|S^K,W,Z^N)\\
&= I(U^N;Y^N)-I(U^N;Z^N)-H(W|S^K,Z^N)-H(U_c|S^K,W,Z^N)\\
&\overset{(15)}{\geq} I(U^N;Y^N)-I(U^N;Z^N)-N[I(\tilde{U};V)-I(\tilde{U};\tilde{Z})+\epsilon_{UV}+\epsilon_{UZ}]\\
&\quad -H(U_c|S^K,W,Z^N)\\
&= N[I(U;Y)-I(U;Z)]-N[I(\tilde{U};V)-I(\tilde{U};\tilde{Z})+\epsilon_{UV}+\epsilon_{UZ}]\\
&\quad -H(U_c|S^K,W,Z^N)\\
&= N[I(U;Y)-I(U;Z)-I(\tilde{U};V)+I(\tilde{U};\tilde{Z})-\epsilon_{UV}-\epsilon_{UZ}]\\
&\quad -H(U_c|S^K,W,Z^N),
\end{aligned}$$

where (12) follows from the fact that $H(S^K,W|U_c,Z^N)=0$; (13) follows from the fact that $H(U_c|Y^N)\geq 0$; (14) follows from the fact that $H(Z^N|U_c)=H(Z^N|U^N)$ and $H(Y^N|U_c)=H(Y^N|U^N)$. (15) follows since the number of subbins in a bin is $2^{N[I(\tilde{U};V)-I(\tilde{U};\tilde{Z})+\epsilon_{UV}+\epsilon_{UZ}]}$.

In step 2, we note that $R_{cam}(\alpha)=I(\tilde{U};\tilde{Y})-I(\tilde{U};V)-\epsilon$ and $\Delta I(\alpha)=I(\tilde{U};\tilde{Z})-I(\tilde{U};V)$. Then, $R_{cam}(\alpha)-\Delta I(\alpha)=I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})-\epsilon$. Furthermore, for $\delta=\epsilon=\epsilon_{UV}=\epsilon_{UZ}=0$,

$$\begin{aligned}
I(U;Y)-I(U;Z) &= I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})\\
R_{cam}(\alpha)(1-\epsilon/2)+\epsilon_{UV}+\epsilon_{UZ}-I(\tilde{U};\tilde{Z})+I(\tilde{U};V) &= R_{cam}(\alpha)-\Delta I(\alpha)\\
&= I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z}).
\end{aligned}$$

Function $R_{cam}(\alpha)(1-\epsilon/2)=(I(\tilde{U};\tilde{Y})-I(\tilde{U};V)-\epsilon)(1-\epsilon/2)$ is parabolic in ϵ, with the minimum at $\epsilon=1+[I(\tilde{U};\tilde{Y})-I(\tilde{U};V)]/2$, and $R_{cam}(\alpha)(1-\epsilon/2)=I(\tilde{U};\tilde{Y})-I(\tilde{U};V)$ when $\epsilon=0$ or $2+I(\tilde{U};\tilde{Y})-I(\tilde{U};V)$. Hence, $R_{cam}(\alpha)(1-\epsilon/2)-\Delta I(\alpha)\leq I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})$ when $\epsilon\in[0,2+I(\tilde{U};\tilde{Y})-I(\tilde{U};V)]$.

For a given $\epsilon\in(0,2+I(\tilde{U};\tilde{Y})-I(\tilde{U};V))$, impose the condition $\epsilon_{UV}+\epsilon_{UZ}<[I(\tilde{U};\tilde{Y})-I(\tilde{U};V)]-R_{cam}(\alpha)(1-\epsilon/2)$ to obtain $R_{cam}(\alpha)(1-\epsilon/2)+\epsilon_{UV}+\epsilon_{UZ}-I(\tilde{U};\tilde{Z})+I(\tilde{U};V)<I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})$.

Now,

$$\begin{aligned}
&I(U;Y) - I(U;Z) \\
&\quad = \frac{1}{2}\log\left[\frac{(P'+\alpha^2Q)(P'+Q+N_1)}{(P'+\alpha^2Q)(P'+Q+N_1)-(P'+\alpha Q)^2}\right] \\
&\quad -\frac{1}{2}\log\left[\frac{(P'+\alpha^2Q)(P'+Q+N_1+N_2)}{(P'+\alpha^2Q)(P'+Q+N_1+N_2)-(P'+\alpha Q)^2}\right] \\
&\quad = \frac{1}{2}\log\left[\frac{(P'+Q+N_1)\{(P'+\alpha^2Q)(P'+Q+N_1+N_2)-(P'+\alpha Q)^2\}}{(P+Q+N_1+N_2)\{(P'+\alpha^2Q)(P'+Q+N_1)-(P'+\alpha Q)^2\}}\right].
\end{aligned}$$

Since the equations

$$\begin{aligned}
(P\phi(\delta)+\alpha^2Q)(P\phi(\delta)+Q+N_1+N_2)-(P\phi(\delta)+\alpha Q)^2 &= 0 \\
(P\phi(\delta)+\alpha^2Q)(P\phi(\delta)+Q+N_1)-(P\phi(\delta)+\alpha Q)^2 &= 0
\end{aligned}$$

have no real solutions for α when $P, Q, N_1, N_2, \delta > 0$, $I(U;Y)-I(U;Z)$ is continuous in δ with the value of $I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})$ at $\delta = 0$, and $R_{cam}(\alpha)(1-\epsilon/2)+\epsilon_{UV}+\epsilon_{UZ}-I(\tilde{U};\tilde{Z})+I(\tilde{U};V)$ is continuous and constant in δ with the value less than $I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})$. If the two curves do not intersect, then there is no constraint on δ. However, if the two curves intersect, then $0 < \delta < \delta^*$, where δ^* is the smallest value such that the two curves intersect. Then the condition on δ becomes $0 < \delta < \delta_2$, where $\delta_2 = \min\{\delta^*, \infty\}$. Thus,

$$I(U;Y) - I(U;Z) - I(\tilde{U};V) + I(\tilde{U};\tilde{Z}) - \epsilon_{UV} - \epsilon_{UZ} \quad \geq \quad R_{cam}(\alpha)(1-\epsilon/2),$$

when

$$\begin{aligned}
\epsilon &\in (0, 2 + I(\tilde{U};\tilde{Y}) - I(\tilde{U};V)), \\
\epsilon_{UV}+\epsilon_{UZ} &< [I(\tilde{U};\tilde{Y}) - I(\tilde{U};V)] - R_{cam}(\alpha)(1-\epsilon/2), \\
\delta &< \delta_2.
\end{aligned}$$

Note that for $\epsilon \geq 2 + I(\tilde{U};\tilde{Y}) - I(\tilde{U};V)$, $I(\tilde{U};\tilde{Y}) - I(\tilde{U};V) - \epsilon < 0$, $1-\epsilon < 0$, $\epsilon > 1$, and there is nothing to prove.

In step 3, the wiretapper's decoding process is considered. In the wiretapper's decoding process, the wiretapper must select a parameter ξ for the joint typical decoder. If $\xi \neq \delta$, the set $T_U^N(\delta) \neq T_U^N(\xi)$, and the wiretapper's decoder is not reliable. Hence, the wiretapper should select $\xi = \delta$, and that is what we will assume in the following analysis.

Firstly, the entropy of the codeword conditioned on bin j, subbin w and the wiretapper's observation is related to the probability of error P_{SB} in the wiretapper's subbin-decoding process through the random variable χ defined as

$$\chi \quad = \quad \begin{cases} 1 & \text{if } \psi(Z^N) \neq U_c \\ 0 & \text{if } \psi(Z^N) = U_c \end{cases},$$

where $\psi(Z^N)$ is the wiretapper's decoding function. Then, the probability of error in the wiretapper's subbin-decoding process is bounded using the union bound, resulting in constraints on δ and N.

Now, the entropy of the codeword conditioned on bin j, subbin w and the wiretapper's observation Z^N is related to the probability of error in the subbin-decoding process.

$$
\begin{aligned}
H(U_c|S^K,W,Z^N) &= H(\chi,U_c|S^K,W,Z^N) - H(\chi|U_c,S^K,W,Z^N) \\
&= H(\chi|S^K,W,Z^N) + H(U_c|\chi,S^K,W,Z^N) - H(\chi|U_c,S^K,W,Z^N) \\
&\overset{(16)}{\leq} H(\chi|S^K,W) + H(U_c|\chi,S^K,W,Z^N) \\
&= \sum_{j=1}^{2^{NR_{cam}(\alpha)}} \sum_{w=1}^{2^{N[I(\tilde{U};V)-I(\tilde{U};\tilde{Z})+\epsilon_{UV}+\epsilon_{UZ}]}} \Pr\{S^K=j,W=w\} \\
&\quad [H(\chi|S^K=j,W=w) \\
&\quad +p(\chi=0|S^K=j,W=w)H(U_c|Z^N,S^K=j,W=w,\chi=0) \\
&\quad +p(\chi=1|S^K=j,W=w)H(U_c|Z^N,S^K=j,W=w,\chi=1)] \\
&= \sum_{j=1}^{2^{NR_{cam}(\alpha)}} \sum_{w=1}^{2^{N[I(\tilde{U};V)-I(\tilde{U};\tilde{Z})+\epsilon_{UV}+\epsilon_{UZ}]}} \Pr\{S^K=j,W=w\}[h(P_{SB}) \\
&\quad +(1-P_{SB})H(U_c|Z^N,S^K=j,W=w,\chi=0) \\
&\quad +P_{SB}H(U_c|Z^N,S^K=j,W=w,\chi=1)] \\
&\overset{(17)}{\leq} \sum_{j=1}^{2^{NR_{cam}(\alpha)}} \sum_{w=1}^{2^{N[I(\tilde{U};V)-I(\tilde{U};\tilde{Z})+\epsilon_{UV}+\epsilon_{UZ}]}} \Pr\{S^K=j,W=w\}[h(P_{SB}) \\
&\quad +P_{SB}\log(2^{N[I(\tilde{U};\tilde{Z})-\epsilon_{UZ}]}-1)] \\
&\overset{(18)}{=} h(P_{SB}) + P_{SB}\log(2^{N[I(\tilde{U};\tilde{Z})-\epsilon_{UZ}]}-1) \\
&\leq h(P_{SB}) + P_{SB}N[I(\tilde{U};\tilde{Z})-\epsilon_{UZ}] \\
&\leq h(P_{SB}) + P_{SB}NI(\tilde{U};\tilde{Z}) \\
\frac{H(U_c|S^K,W,Z^N)}{R_{cam}(\alpha)N} &\leq \frac{h(P_{SB}) + P_{SB}NI(\tilde{U};\tilde{Z})}{R_{cam}(\alpha)N} \\
&\leq \epsilon/2,
\end{aligned}
$$

for $N \geq L_3 = 2h(P_{SB})/\{R_{cam}(\alpha)\epsilon - 2P_{SB}I(\tilde{U};\tilde{Z})\}$, and $P_{SB} = p(\chi=1|S^K=j,W=w) < R_{cam}(\alpha)\epsilon/2I(\tilde{U};\tilde{Z})$, where $h(P_{SB})$ is the binary entropy function. (16) follows from the facts that $H(\chi|S^K,W,Z^N) \leq H(\chi|S^K,W)$ and $H(\chi|U_c,S^K,W,Z^N)=0$. (17) follows from the fact that the number of the auxiliary codewords per bin is $2^{N[I(\tilde{U};\tilde{Z})-\epsilon_{UZ}]}$. (18) follows from the fact that P_{SB} is independent of the bin and the subbin.

To bound P_{SB}, define the following events associated with the wiretapper's subbin-decoding process:

- $\mathcal{E}^{Z1}(j,w)$: in the wiretapper's subbin-decoding process, there is no sequence u_c in bin j, subbin w that is jointly typical with the received sequence z^N.
- $\mathcal{E}^{Z2}(j,w)$: in the wiretapper's subbin-decoding process, a sequence $u_c \neq u_c(j,k^*)$ is jointly typical with the received sequence z^N for u_c in bin j, subbin w.

An error in the wiretapper's subbin-decoding process occurs when at least one of the events $\mathcal{E}^{Z1}(j,w)$ and $\mathcal{E}^{Z2}(j,w)$ occurs when there is no encoding error. Then the probability of

wiretapper's subbin-decoding error given the secret message is

$$\begin{aligned} P_{SB} &= p(\chi = 1 | S^K = j, W = w) \\ &\leq \Pr\{\mathcal{E}^{Z1}(j,w) | \mathcal{E}^X(j)^C \mathcal{E}^V(j)^C\} + \Pr\{\mathcal{E}^{Z2}(j) | \mathcal{E}^X(j)^C \mathcal{E}^V(j)^C\} \\ &= \Pr\{\mathcal{E}^{Z1}(1,1) | \mathcal{E}^X(1)^C \mathcal{E}^V(1)^C\} + \Pr\{\mathcal{E}^{Z2}(1,1) | \mathcal{E}^X(1)^C \mathcal{E}^V(1)^C\}. \end{aligned}$$

The probabilities $\Pr\{\mathcal{E}^{Z1}(1,1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}$ and $\Pr\{\mathcal{E}^{Z2}(1,1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}$ are evaluated as follows:

$\Pr\{\mathcal{E}^{Z1}(1,1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}$. By Lemma 2.5 with $\epsilon_1 = R_{cam}(\alpha)\epsilon/6I(\tilde{U};\tilde{Z})$, there exists L_4 such that, for a given $\delta > 0$, $\Pr\{(U_c, Z^N) \in T^N_{U,Z}(\delta)\} > 1 - R_{cam}(\alpha)\epsilon/6I(\tilde{U};\tilde{Z})$ when $N \geq L_4$, which implies that $\Pr\{\mathcal{E}^{Z1}(1,1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\} \leq R_{cam}(\alpha)\epsilon/6I(\tilde{U};\tilde{Z})$ for a given $\delta > 0$ and $N \geq L_4$.

$\Pr\{\mathcal{E}^{Z2}(1,1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}$. By the code generating process, $U_c(1,k^*)$ and $U_c(1,k)$ are independent, and so are Z^N and $U_c(1,k)$ for $k \neq k^*$. By inequality (2.20), $\Pr\{(U_c(1,k), Z^N) \in T^N_{U,Z}(\delta) < 2^{-N[I(U;Z)-3\delta]}$ for $k \neq k^*$. Hence,

$$\begin{aligned} \Pr\{\mathcal{E}^{Z2}(1,1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\} &\leq \sum_{k \neq k^*} 2^{-N[I(U;Z)-3\delta]} \\ &\overset{(19)}{\leq} (2^{N[I(\tilde{U};\tilde{Z})-\epsilon_{UZ}]} - 1) 2^{-N[I(U;Z)-3\delta]} \\ &< 2^{-N[I(U;Z)-3\delta-I(\tilde{U};\tilde{Z})+\epsilon_{UZ}]} \\ &\leq R_{cam}(\alpha)\epsilon/6I(\tilde{U};\tilde{Z}), \end{aligned}$$

when

$$I(U;Z) - I(\tilde{U};\tilde{Z}) - 3\delta + \epsilon_{UZ} > 0,$$

and

$$\begin{aligned} N[I(U;Z) - I(\tilde{U};\tilde{Z}) - 3\delta + \epsilon_{UZ}] &\geq \log\left[\frac{6I(\tilde{U};\tilde{Z})}{R_{cam}(\alpha)\epsilon}\right] \\ N &\geq L_5, \end{aligned}$$

where

$$L_5 = \frac{1}{I(U;Z) - I(\tilde{U};\tilde{Z}) - 3\delta + \epsilon_{UZ}} \log\left[\frac{6I(\tilde{U};\tilde{Z})}{R_{cam}(\alpha)\epsilon}\right].$$

(19) follows from the union bound.

Now, define $\lambda_Z(\delta) = I(U;Z) - I(\tilde{U};\tilde{Z}) - 3\delta + \epsilon_{UZ}$, and $\phi(\delta) = (1 + 4\delta\ln(2))^{-1}$. We then have that

$$\begin{aligned} \lambda_Z(\delta) &\overset{(20)}{=} \frac{1}{2}\log\left[\frac{(P\phi(\delta) + \alpha^2 Q)(P\phi(\delta) + Q + N_1 + N_2)}{(P\phi(\delta) + \alpha^2 Q)(P\phi(\delta) + Q + N_1 + N_2) - (P\phi(\delta) + \alpha Q)^2}\right] \\ &\quad - \frac{1}{2}\log\left[\frac{(P + \alpha^2 Q)(P + Q + N_1 + N_2)}{(P + \alpha^2 Q)(P + Q + N_1 + N_2) - (P + \alpha Q)^2}\right] - 3\delta + \epsilon_{UZ} \\ \lambda_Z(0) &\overset{(21)}{>} 0 \text{ for } 3\delta < \epsilon_{UZ} \qquad (4.12) \\ \lim_{\delta \to \infty} \lambda_Y(\delta) &= -\infty \text{ for finite } \alpha, P, Q, N_1, N_2 \qquad (4.13) \end{aligned}$$

(20) follows from equation (4.4). (21) follows from the fact that $\phi(0) = 1$.

Furthermore, $\lambda_Y(\delta)$ is a continuous function because the equations

$$\begin{aligned}
(P\phi(\delta) + \alpha^2 Q)(P\phi(\delta) + Q + N_1 + N_2) &= 0 \\
(P\phi(\delta) + \alpha^2 Q)(P\phi(\delta) + Q + N_1 + N_2) - (P\phi(\delta) + \alpha Q)^2 &= 0 \\
(P + \alpha^2 Q)(P + Q + N_1 + N_2) &= 0 \\
(P + \alpha^2 Q)(P + Q + N_1 + N_2) - (P + \alpha Q)^2 &= 0
\end{aligned}$$

have no real solutions for α when $P, Q, N_1, N_2, \delta > 0$.

Consequently, there exists the smallest $\delta_3 > 0$ such that $\lambda_Z(\delta_3) = 0$. Thus, $\lambda_Z(\delta) > 0$ for $0 < \delta < \delta_3$ and $3\delta < \epsilon_{UZ}$, and $\Pr\{\mathcal{E}^{Z2}(1,1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\} \leq \epsilon/3$ for $0 < \delta < \min\{\delta_3, \epsilon_{UZ}/3\}$ and $N \geq L_5$.

For the coding strategy to work, given an $\epsilon \in (0, 2 + I(\tilde{U};\tilde{Y}) - I(\tilde{U};V))$, pick $0 < \epsilon_{UV} < \epsilon$ and $\epsilon_{UZ} > 0$ such that $\epsilon_{UV} + \epsilon_{UZ} < [I(\tilde{U};\tilde{Y}) - I(\tilde{U};V)] - R_{cam}(\alpha_{cam})(1 - \epsilon/2)$, pick δ such that $0 < \delta < \min\{1, \delta_0, \delta_1, \delta_2, \delta_3, \epsilon_{UZ}/3\}$, and pick $N \geq N_{min} = \max\{L_0, L_1, L_2, L_3, L_4, L_5\}$.

Finally, we must ensure that $\Delta I(\alpha_{cam}) \leq 0$, where $\Delta I(\alpha_{cam})$

$$\begin{aligned}
&= \frac{1}{2}\log\left[\frac{P(P + Q + N_1 + N_2)}{(P + \alpha_{cam}^2 Q)(P + Q + N_1 + N_2) - (P + \alpha_{cam} Q)^2}\right] \\
&= \frac{1}{2}\log\left[\frac{P(P + Q + N_1 + N_2)(P + N_1)^2}{(P(P + N_1)^2 + P^2 Q)(P + Q + N_1 + N_2) - (P(P + N_1) + PQ)^2}\right] \\
&= \frac{1}{2}\log\left[\frac{(P + Q + N_1 + N_2)(P + N_1)^2}{((P + N_1)^2 + PQ)(P + Q + N_1 + N_2) - P(P + Q + N_1)^2}\right].
\end{aligned}$$

$\Delta I(\alpha_{cam}) \leq 0$ when

$$\begin{aligned}
(P + N_1)^2(P + Q + N_1 + N_2) &\leq ((P + N_1)^2 + PQ)(P + Q + N_1 + N_2) \\
&\quad -P(P + Q + N_1)^2 \\
0 &\leq PQ(P + Q + N_1 + N_2) - P(P + Q + N_1)^2 \\
&= P[PQ + Q^2 + QN_1 + QN_2 \\
&\quad -P^2 - Q^2 - N_1^2 - 2PQ - 2PN_1 - 2QN_1] \\
&\geq P[P^2 + P(Q + 2N_1) + N_1^2 + QN_1 - QN_2]
\end{aligned}$$

which is true when

$$0 \leq \quad P \quad \leq -N_1 - \frac{Q}{2} + \frac{\sqrt{Q^2 + 4QN_2}}{2}.$$

Finally, the upper bound on P is greater than 0 when

$$\begin{aligned} -N_1 - \frac{Q}{2} + \frac{1}{2}\sqrt{Q^2 + 4QN_2} &> 0 \\ Q^2 + 4QN_2 &> (2N_1 + Q)^2 \\ N_2 &> \frac{4N_1^2 + 4QN_1}{4Q} \\ &= N_1 + \frac{N_1^2}{Q}. \end{aligned}$$

■

The system operating in the camouflage mode at the rate R_{cam}, with $d \geq 1 - \epsilon$ and $P_e \leq \epsilon$ is *optimal* since the parameter α maximizes the rate while keeping the condition on the equivocation and the probability of error satisfied. However, it cannot always be used due to the condition on the leakage function. Next, the efficient camouflage mode is defined as an alternative to the optimal camouflage mode.

The Efficient Camouflage Mode

We say that the system operates in the camouflage mode *efficiently* if the average information leakage rate is zero, i.e., $\Delta I(\alpha) = 0$ or $I(\tilde{U};\tilde{Z}) = I(\tilde{U};V)$. Given the system parameters P, Q, N_1, and N_2, the system operates efficiently if

$$\alpha = \alpha_0 = \frac{P}{P + N_1 + N_2}\left[1 + \sqrt{\frac{P + Q + N_1 + N_2}{Q}}\right]$$

since, from equation (4.3), $I(\tilde{U};\tilde{Z}) = I(\tilde{U};V)$ when

$$\begin{aligned} (P + \alpha^2 Q)(P + Q + N_1 + N_2) - (P + \alpha Q)^2 &= P(P + Q + N_1 + N_2) \\ \alpha^2 Q(P + Q + N_1 + N_2) &= P^2 + 2\alpha PQ + \alpha^2 Q^2 \\ \alpha^2 Q(P + N_1 + N_2) &= P^2 + 2\alpha PQ \\ 0 &= \alpha^2 Q(P + N_1 + N_2) - 2\alpha PQ - P^2. \end{aligned}$$

That is

$$\alpha = \alpha_0 = \frac{P}{P + N_1 + N_2}\left[1 + \sqrt{\frac{P + Q + N_1 + N_2}{Q}}\right].$$

Hence, the efficient camouflage rate for the system with parameters $P, Q, N_1, N_2, \alpha = \alpha_0$ is

$$R_{cam}(\alpha_0) = \frac{1}{2}\log\left[\frac{P(P + Q + N_1)}{(P + \alpha_0^2 Q)(P + Q + N_1) - (P + \alpha_0 Q)^2}\right]$$

The system can always operate in the efficient camouflage mode for non-zero α while the optimal camouflage mode can be implemented for restricted P and N_2. The efficient camouflage mode is not optimal in term of rate of transmission but ensures that there is no leaking of the secret message to the wiretapper.

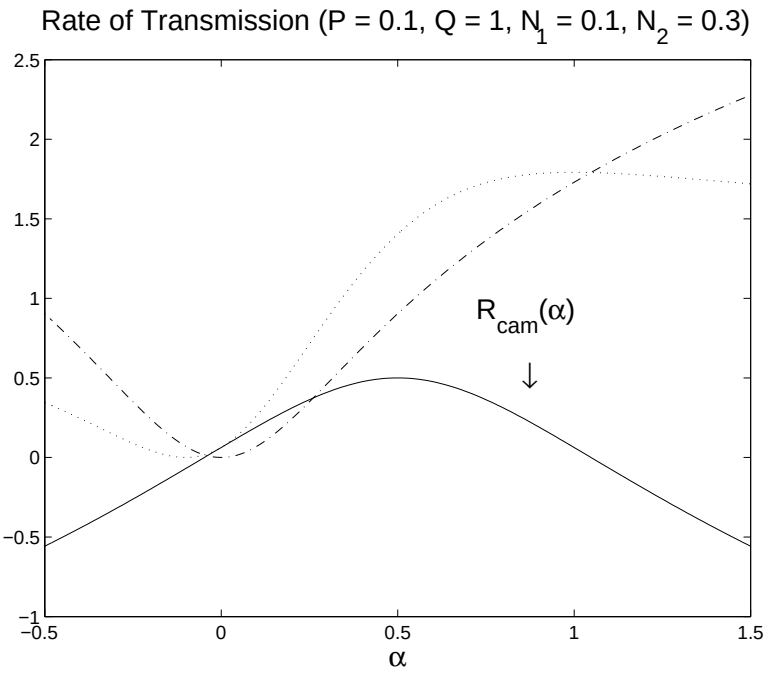

Figure 4.5: The rate of transmission as a function of α for the camouflage mode.

4.3.4 High-power Mode

Since the camouflage mode only works when $\Delta I(\alpha) \leq 0$, we define the high-power mode for the case $\Delta I(\alpha) > 0$ and prove an achievable rate at perfect secrecy with this mode of operation. Given an $\epsilon > 0$, let ϵ_{UZ} be a constant associated with the codebook generating process, α

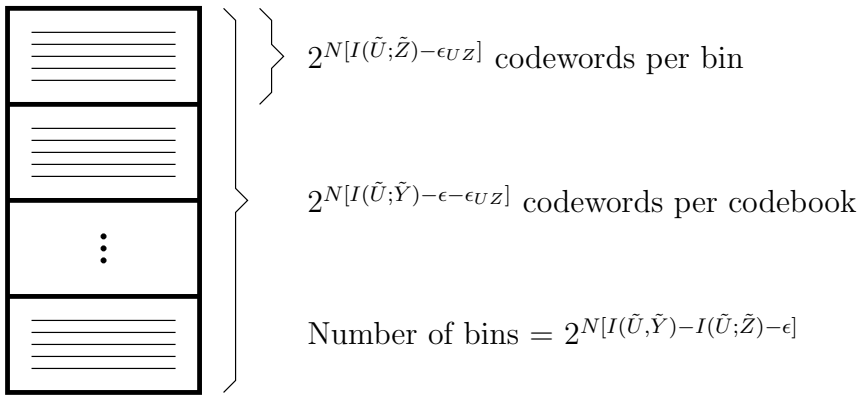

Figure 4.6: The codebook for the high-power mode.

and δ be constants associated with the encoding and decoding processes, and ξ be a constant associated with the wiretappers's decoding process to be specified later. For $\Delta I(\alpha) > 0$,

1. *Generating the codebook.* Generate $2^{N[I(\tilde{U};\tilde{Y})-\epsilon-\epsilon_{UZ}]}$ sequences u^N according to the dis-

tribution $p(u^N) = \prod_{i=1}^{N} p(u_i)$, and $p(u_i) \sim \mathcal{N}(0, P' + \alpha^2 Q)$ for all $i \in \{1, 2, \ldots, N\}$. Place the sequences u^N randomly into $2^{N[I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})-\epsilon]}$ bins so that each bin contains $2^{N[I(\tilde{U};\tilde{Z})-\epsilon_{UZ}]}$ sequences. Index each bin by $j \in \{1, 2, \ldots, 2^{NR}\}$, where $R = I(\tilde{U};\tilde{Y}) - I(\tilde{U};\tilde{Z}) - \epsilon$. The codebook is given to the sender and the recipient.

2. *Encoding.* To send a message j through an interference v^N, the sender looks for a sequence u_c in bin j such that $(u_c, v^N) \in T^N_{U,V}(\delta)$ and transmit $x^N = u_c - \alpha v^N$. If there is more than one sequence u_c that is jointly typical with v^N, randomly select one.

3. *Decoding.* To decode the received sequence for the message, the recipient finds a sequence u_c in the codebook that is jointly typical with the received sequence y^N, i.e., $(u_c, y^N) \in T^N_{U,Y}(\delta)$. Declare the index to the bin, in which the sequence is found as the received message.

4. *Wiretapper's decoding.* The wiretapper receives a sequence z^N and finds a sequence u_c in the codebook that is jointly typical with the received sequence z^N, i.e., $(u_c, z^N) \in T^N_{U,Z}(\xi)$. Declare the index to the bin, in which the sequence is found as the received message.

5. *Probability of error.* An error occurs in the coding process when a message j is to be transmitted and one or more of the following events occurs.

   - $\mathcal{E}^V(j)$: in the encoding process, there is no sequence u_c in bin j that is jointly typical with the interference sequence.
   - $\mathcal{E}^X(j)$: in the encoding process, $x^N = u_c - \alpha v^N$ does not satisfy the power constraint.
   - $\mathcal{E}^{Y1}(j)$: in the decoding process, there is no sequence u_c that is jointly typical with the received sequence.
   - $\mathcal{E}^{Y2}(j)$: in the decoding process, a sequence u_c in bin $i \neq j$ is jointly typical with the received sequence.

Theorem 4.3 *Given the high-power mode of operation to be used in the Gaussian wiretap channel with side information and average power constraint P, for any $\epsilon > 0$ and*

$$P > -\frac{Q}{2} + \frac{\sqrt{Q^2 + 4Q(N_1 + N_2)}}{2},$$

there exists an integer N_{min} such that, for $N \geq N_{min}$,

$$\begin{aligned} \frac{H(S^K)}{N} &\geq R_{high} \triangleq \frac{1}{2} \log \left[\frac{(P + Q + N_1)(N_1 + N_2)}{(P + Q + N_1 + N_2)N_1} \right] - \epsilon \\ \frac{H(S^K|Z^N)}{H(S^K)} &\geq 1 - \epsilon \\ P_e &\leq \epsilon. \end{aligned}$$

Proof: This proof consists of three main parts: proof of the rate, proof of the probability of error, and proof of the equivocation.

Proof of the Rate. The number of messages that can be sent is the number of bins in the codebook, $2^{NR_{high}(\alpha)} = 2^{N[I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})-\epsilon]}$. Consequently, $R_{high}(\alpha)$

$$\begin{aligned}
&= [I(\tilde{U};\tilde{Y}) - I(\tilde{U};\tilde{Z}) - \epsilon] \\
&\overset{(1)}{=} \frac{1}{2}\log\left[\frac{(P+\alpha^2Q)(P+Q+N_1)}{(P+\alpha^2Q)(P+Q+N_1)-(P+\alpha Q)^2}\right] \\
&\quad -\frac{1}{2}\log\left[\frac{(P+\alpha^2Q)(P+Q+N_1+N_2)}{(P+\alpha^2Q)(P+Q+N_1+N_2)-(P+\alpha Q)^2}\right] - \epsilon \\
&= \frac{1}{2}\log\left[\frac{(P+Q+N_1)\{(P+\alpha^2Q)(P+Q+N_1+N_2)-(P+\alpha Q)^2\}}{\{(P+\alpha^2Q)(P+Q+N_1)-(P+\alpha Q)^2\}(P+Q+N_1+N_2)}\right] - \epsilon
\end{aligned}$$

(1) follows from equations (4.1) and (4.4). We select the optimal α for this coding strategy by setting $dR_{high}(\alpha)/d\alpha$ to 0, resulting in

$$\begin{aligned}
0 &= [\{(P+\alpha^2Q)(P+Q+N_1)-(P+\alpha Q)^2\}(P+Q+N_1+N_2)] \times \\
&\quad [(P+Q+N_1)\{2\alpha Q(P+Q+N_1+N_2)-2Q(P+\alpha Q)\}] \\
&\quad -[(P+Q+N_1)\{(P+\alpha^2Q)(P+Q+N_1+N_2)-(P+\alpha Q)^2\}] \times \\
&\quad [\{2\alpha Q(P+Q+N_1)-2Q(P+\alpha Q)\}(P+Q+N_1+N_2)] \\
&= [\{(P+\alpha^2Q)(P+Q+N_1)-(P+\alpha Q)^2\}] \times \\
&\quad [\{2\alpha Q(P+Q+N_1+N_2)-2Q(P+\alpha Q)\}] \\
&\quad -[\{(P+\alpha^2Q)(P+Q+N_1+N_2)-(P+\alpha Q)^2\}] \times \\
&\quad [\{2\alpha Q(P+Q+N_1)-2Q(P+\alpha Q)\}] \\
&= [(P+\alpha^2Q)(P+Q+N_1)\{-2Q(P+\alpha Q)\} \\
&\quad -(P+\alpha Q)^2\{2\alpha Q(P+Q+N_1+N_2)\}] \\
&\quad -[(P+\alpha^2Q)(P+Q+N_1+N_2)\{-2Q(P+\alpha Q)\} \\
&\quad -(P+\alpha Q)^2\{2\alpha Q(P+Q+N_1)\}] \\
&= N_2(P+\alpha^2Q)\{2Q(P+\alpha Q)\} - N_2(P+\alpha Q)^2(2\alpha Q) \\
&= [2N_2Q(P+\alpha Q)] \times [(P+\alpha^2Q) - \alpha(P+\alpha Q)] \\
&= (P+\alpha Q)(P-\alpha P).
\end{aligned}$$

Hence, $dR_{high}(\alpha)/d\alpha = 0$ when $\alpha = 1, -P/Q$. $R_{high}(\alpha)$ reaches its minimum value at $\alpha = -P/Q$ and maximum value at $\alpha = \alpha_{high} = 1$, and the maximum value is

$$\begin{aligned}
R_{high}(1) &= \frac{1}{2}\log\left[\frac{(P+Q+N_1)\{(P+Q)(P+Q+N_1+N_2)-(P+Q)^2\}}{\{(P+Q)(P+Q+N_1)-(P+Q)^2\}(P+Q+N_1+N_2)}\right] - \epsilon \\
&= \frac{1}{2}\log\left[\frac{(P+Q+N_1)\{(P+Q+N_1+N_2)-(P+Q)\}}{\{(P+Q+N_1)-(P+Q)\}(P+Q+N_1+N_2)}\right] - \epsilon \\
&= \frac{1}{2}\log\left[\frac{(P+Q+N_1)(N_1+N_2)}{N_1(P+Q+N_1+N_2)}\right] - \epsilon.
\end{aligned}$$

Proof of the probability of error. The probability of error is bounded by the probabilities of $\{\mathcal{E}^V(1)\}$ $\{\mathcal{E}^X(1)|\mathcal{E}^V(1)^C\}$, $\{\mathcal{E}^{Y1}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}$, and $\{\mathcal{E}^{Y2}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}$, each of

which can be bounded in the following steps.

Denote the k^{th} auxiliary codeword in bin j by $u_c(j,k)$, where $k \in \{1,2,\ldots,2^{N[I(\tilde{U};\tilde{Z})-\epsilon_{UZ}]}\}$ and the selected auxiliary codeword in bin j, which is jointly typical with the interference, by $u_c(j,k^*)$. The transmitted codeword is then $x^N = u_c(j,k^*) - \alpha v^N$.

$$\begin{aligned} P_e &\overset{(2)}{\leq} \sum_{j=1}^{2^{NR_{high}(\alpha)}} \Pr\{S^K = j\}[\Pr\{\mathcal{E}^V(j)\} + \Pr\{\mathcal{E}^X(j)|\mathcal{E}^V(j)^C\} \\ &\quad + \Pr\{\mathcal{E}^{Y1}(j)|\mathcal{E}^X(j)^C\mathcal{E}^V(j)^C\} + \Pr\{\mathcal{E}^{Y2}(j)|\mathcal{E}^X(j)^C\mathcal{E}^V(j)^C\}] \\ &\overset{(3)}{=} \Pr\{\mathcal{E}^V(1)\} + \Pr\{\mathcal{E}^X(1)|\mathcal{E}^V(1)^C\} \\ &\quad + \Pr\{\mathcal{E}^{Y1}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\} + \Pr\{\mathcal{E}^{Y2}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}]. \end{aligned}$$

(2) follows from the union bound. (3) follows from the fact that error events don't depend on the message j. We now consider the four terms in the summation separately.

$\Pr\{\mathcal{E}^V(1)\}$. By the code generation process, U_c and V^N are independent. By equation (2.21),

$$\begin{aligned} \Pr\{(U_c,V^N) \notin T^N_{U,V}(\delta)\} &\leq 1-(1-\delta)2^{-N[I(U;V)+3\delta]} \\ \Pr\{\mathcal{E}^V(1)\} &\overset{(4)}{\leq} [1-(1-\delta)2^{-N[I(U;V)+3\delta]}]^{2^{N[I(\tilde{U};\tilde{Z})-\epsilon_{UZ}]}} \\ &\overset{(5)}{\leq} \exp\{-(1-\delta)2^{-N[I(U;V)+3\delta]}\}^{2^{N[I(\tilde{U};\tilde{Z})-\epsilon_{UZ}]}} \\ &= \exp\{-(1-\delta)2^{N[I(\tilde{U};\tilde{Z})-I(U;V)-\epsilon_{UZ}-3\delta]}\} \\ &\leq \epsilon/3 \end{aligned}$$

when $\delta < 1$, and

$$I(\tilde{U};\tilde{Z}) - I(U;V) - \epsilon_{UZ} - 3\delta > 0 \tag{4.14}$$

and

$$\begin{aligned} (1-\delta)2^{N[I(\tilde{U};\tilde{Z})-I(U;V)-\epsilon_{UZ}-3\delta]} &\geq \ln\left[\frac{3}{\epsilon}\right] \\ N[I(\tilde{U};\tilde{Z}) - I(U;V) - \epsilon_{UZ} - 3\delta] &\geq \log\left[\frac{1}{1-\delta}\ln\left[\frac{3}{\epsilon}\right]\right] \\ N &\geq L_0, \end{aligned}$$

where

$$L_0 = \frac{1}{I(\tilde{U};\tilde{Z}) - I(U;V) - \epsilon_{UZ} - 3\delta}\log\left[\frac{1}{1-\delta}\ln\left[\frac{3}{\epsilon}\right]\right].$$

(4) follows from the fact that there are $2^{N[I(\tilde{U};\tilde{Z})-\epsilon_{UZ}]}$ codewords in a bin. (5) follows from the fact that $e^a \geq 1+a$ (Taylor's expansion of e^a) [18].

We now consider the condition when inequality (4.14) is satisfied as follows.

$$\begin{aligned} I(\tilde{U};\tilde{Z}) - I(U;V) - \epsilon_{UZ} - 3\delta &= I(\tilde{U};\tilde{Z}) + [-I(\tilde{U};V) + I(\tilde{U};V)] - I(U;V) - \epsilon_{UZ} - 3\delta \\ &= \Delta I(\alpha) + I(\tilde{U};V) - I(U;V) - \epsilon_{UZ} - 3\delta \end{aligned}$$

Let $g_{LHS}(\delta) = I(\tilde{U};V) - I(U;V)$, and $g_{RHS}(\delta) = 3\delta + \epsilon_{UZ} - \Delta I(\alpha)$. Then, $g_{LHS}(\delta)$ and $g_{RHS}(\delta)$ are continuous functions. (See the proof of Theorem 4.2.)

Imposing the condition $\Delta I(\alpha) > \epsilon_{UZ}$, we have $g_{RHS}(0) < 0$. From the facts that $g_{LHS}(0) = 0$, $g_{RHS}(0) = \epsilon_{UZ} - \Delta I(\alpha) < 0$, and $g_{LHS}(\delta)$ is a monotonically decreasing continuous function approaching $-\infty$ while $g_{RHS}(\delta)$ is a monotonically increasing continuous function approaching ∞ as δ increases, it follows that there exists a $\delta_0 > 0$ such that $g_{LHS}(\delta_0) = g_{RHS}(\delta_0)$ by the continuity of the functions [13]. Thus, $g_{LHS}(\delta) - g_{RHS}(\delta) = I(\tilde{U};V) - I(U;V) - \epsilon_{UZ} - 3\delta + \Delta I(\alpha) > 0$ for $0 < \delta < \delta_0$ and $\epsilon_{UZ} < \Delta I(\alpha)$, and $\Pr\{\mathcal{E}^V(1)\} \leq \epsilon/3$ when $0 < \delta < \min\{1, \delta_0\}$, $\epsilon_{UZ} < \Delta I(\alpha)$ and $N \geq L_0$.

$\Pr\{\mathcal{E}^X(1)|\mathcal{E}^V(1)^C\}$. If the event $\mathcal{E}^V(1)$ does not occur, then we have that $(u_c(1,k^*), v^N) \in T^N_{U,V}(\delta)$. By Lemma 2.12, $\frac{1}{N}\sum_{i=1}^N x_i^2 \leq P$, and $\Pr\{\mathcal{E}^X(1)|\mathcal{E}^V(1)^C\} = 0$.

$\Pr\{\mathcal{E}^{Y1}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}$. By Lemma 2.5 with $\epsilon_1 = \epsilon/3$, there exists an integer L_1 for a given $\delta > 0$ such that $\Pr\{(U_c, Y^N) \in T^N_{U,Y}(\delta)\} > 1 - \epsilon/3$ when $N \geq L_1$, which implies that $\Pr\{\mathcal{E}^{Y1}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\} \leq \epsilon/3$ for a given $\delta > 0$ and $N \geq L_1$.

$\Pr\{\mathcal{E}^{Y2}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}$. By the code generating process, $U_c(1,k^*)$ and $U_c(1,k)$ are independent, and so are Y^N and $U_c(1,k)$ for $(i,k) \neq (1,k^*)$. By inequality (2.20), $\Pr\{(U_c(i,k), Y^N) \in T^N_{U,Y}(\delta) < 2^{-N[I(U;Y)-3\delta]}$ for $(i,k) \neq (1,k^*)$. Hence,

$$\begin{aligned}
\Pr\{\mathcal{E}^{Y2}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}\} &\leq \sum_{(i,k)\neq(1,k^*)} 2^{-N[I(U;Y)-3\delta]} \\
&\overset{(6)}{\leq} (2^{N[I(\tilde{U};\tilde{Y})-\epsilon-\epsilon_{UZ}]} - 1)2^{-N[I(U;Y)-3\delta]} \\
&< 2^{-N[I(U;Y)-3\delta-I(\tilde{U};\tilde{Y})+\epsilon+\epsilon_{UZ}]} \\
&\leq \epsilon/3
\end{aligned}$$

when

$$I(U;Y) - I(\tilde{U};\tilde{Y}) - 3\delta + \epsilon + \epsilon_{UZ} > 0,$$

and

$$\begin{aligned}
N[I(U;Y) - I(\tilde{U};\tilde{Y}) - 3\delta + \epsilon + \epsilon_{UZ}] &\geq \log\left[\frac{3}{\epsilon}\right] \\
N &\geq L_2,
\end{aligned}$$

where

$$L_2 = \frac{1}{I(U;Y) - I(\tilde{U};\tilde{Y}) - 3\delta + \epsilon + \epsilon_{UZ}} \log\left[\frac{3}{\epsilon}\right].$$

(6) follows from the union bound.

Now, define $\lambda_Y(\delta) = I(U;Y) - I(\tilde{U};\tilde{Y}) - 3\delta + \epsilon + \epsilon_{UZ}$, and $\phi(\delta) = (1 + 4\delta \ln(2))^{-1}$. We then have that

$$\begin{aligned}
\lambda_Y(\delta) &\overset{(7)}{=} \frac{1}{2}\log\left[\frac{(P\phi(\delta)+\alpha^2 Q)(P\phi(\delta)+Q+N_1)}{(P\phi(\delta)+\alpha^2 Q)(P\phi(\delta)+Q+N_1)-(P\phi(\delta)+\alpha Q)^2}\right] \\
&\quad -\frac{1}{2}\log\left[\frac{(P+\alpha^2 Q)(P+Q+N_1)}{(P+\alpha^2 Q)(P+Q+N_1)-(P+\alpha Q)^2}\right] - 3\delta + \epsilon + \epsilon_{UZ} \\
\lambda_Y(0) &\overset{(8)}{>} 0 \qquad (4.15) \\
\lim_{\delta\to\infty} \lambda_Y(\delta) &= -\infty \text{ for finite } \alpha, P, Q, N_1 \qquad (4.16)
\end{aligned}$$

(7) follows from equation (4.4). (8) follows from the fact that $\phi(0) = 1$.

Since $\lambda_Y(\delta)$ is a continuous function (See the proof in Theorem 4.2), there exists the smallest $\delta_1 > 0$ such that $\lambda_Y(\delta_1) = 0$. Thus, $\lambda_Y(\delta) = I(U;Y) - I(\tilde{U};\tilde{Y}) - 3\delta + \epsilon + \epsilon_{UZ} > 0$ for $0 < \delta < \delta_1$, and $\Pr\{\mathcal{E}^{Y^2}(1)|\mathcal{E}^X(1)^C\} \leq \epsilon/3$ for $0 < \delta < \delta_1$ and $N \geq L_2$.

Combining the four terms of the summation,

$$P_e \leq \epsilon/3 + 0 + \epsilon/3 + \epsilon/3 = \epsilon,$$

for $0 < \delta < \min\{1, \delta_0, \delta_1\}$, $0 < \epsilon_{UZ} < \Delta I(\alpha)$, and $N \geq \max\{L_0, L_1, L_2\}$.

Proof of the equivocation. This part of the proof is done in three steps:

1. show that $H(S^K|Z^N) = N[I(U;Y) - I(U;Z)] - H(U_c|S^K, Z^N)$;
2. show that $I(U;Y) - I(U;Z) \geq R_{high}(\alpha)(1 - \epsilon/2)$;
3. show that $H(U_c|S^K, Z^N)/R_{high}(\alpha)N \leq \epsilon/2$.

Combining the above three steps,

$$\begin{aligned}
\frac{H(S^K|Z^N)}{H(S^K)} &= \frac{N[I(U;Y) - I(U;Z)] - H(U_c|S^K, Z^N)}{R_{high}(\alpha)N} \\
&\geq 1 - \epsilon/2 - \frac{H(U_c|S^K, Z^N)}{R_{high}(\alpha)N} \\
&\geq 1 - \epsilon/2 - \epsilon/2 \\
&= 1 - \epsilon.
\end{aligned}$$

Proceeding to the first step,

$$\begin{aligned}
H(S^K|Z^N) &= H(S^K,Z^N)-H(Z^N)\\
&= H(S^K,U_c,Z^N)-H(U_c|S^K,Z^N)-H(Z^N)\\
&= H(U_c,Z^N)+H(S^K|U_c,Z^N)-H(U_c|S^K,Z^N)-H(Z^N)\\
&\overset{(9)}{=} H(U_c|Z^N)-H(U_c|S^K,Z^N)\\
&\overset{(10)}{\geq} H(U_c|Z^N)-H(U_c|Y^N)-H(U_c|S^K,Z^N)\\
&= H(U_c,Z^N)-H(Z^N)-H(U_c,Y^N)+H(Y^N)-H(U_c|S^K,Z^N)\\
&= [H(U_c)+H(Z^N|U_c)]-H(Z^N)-[H(U_c)+H(Y^N|U_c)]\\
&\quad +H(Y^N)-H(U_c|S^K,Z^N)\\
&= H(Z^N|U_c)-H(Z^N)-H(Y^N|U_c)+H(Y^N)-H(U_c|S^K,Z^N)\\
&= H(Z^N|U_c)-[H(U^N,Z^N)-H(U^N|Z^N)]-H(Y^N|U_c)\\
&\quad +[H(U^N,Y^N)-H(U^N|Y^N)]-H(U_c|S^K,Z^N)\\
&= H(Z^N|U_c)-[H(U^N)+H(Z^N|U^N)-H(U^N|Z^N)]-H(Y^N|U_c)\\
&\quad +[H(U^N)+H(Y^N|U^N)-H(U^N|Y^N)]-H(U_c|S^K,Z^N)\\
&\overset{(11)}{=} H(U^N|Z^N)-H(U^N|Y^N)-H(U_c|S^K,Z^N)\\
&= I(U^N;Y^N)-I(U^N;Z^N)-H(U_c|S^K,Z^N)\\
&= N[I(U;Y)-I(U;Z)]-H(U_c|S^K,Z^N)
\end{aligned}$$

where (9) follows from the fact that $H(S^K|U_c,Z^N)=0$. (10) follows from the fact that $H(U_c|Y^N)\geq 0$. (11) follows from the fact that $H(Z^N|U_c)=H(Z^N|U^N)$ and $H(Y^N|U_c)=H(Y^N|U^N)$ since the realization of U_c is generated from U^N.

In step 2, we note that the function $R_{high}(\alpha)(1-\epsilon/2)=(I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})-\epsilon)(1-\epsilon/2)$ is parabolic in ϵ, with the minimum at $\epsilon=1+[I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})]/2$, and $R_{high}(\alpha)(1-\epsilon/2)=I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})$ when $\epsilon=0$ or $2+I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})$. Hence, $R_{high}(\alpha)(1-\epsilon/2)<I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})$ when $\epsilon\in(0,2+I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z}))$.

Now, $I(U;Y)-I(U;Z)$ is continuous in δ with the value of $I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})$ at $\delta=0$, and $R_{high}(\alpha)(1-\epsilon/2)$ is continuous and constant in δ with the value less than $I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})$. If the two curves do not intersect, then there is no constraint on δ. However, if the two curves intersect, then $0<\delta<\delta^*$, where δ^* is the smallest value such that the two curves intersect. Then the condition on δ becomes $0<\delta<\delta_2$, where $\delta_2=\min\{\delta^*,\infty\}$. Thus,

$$I(U;Y)-I(U;Z) \quad\geq\quad R_{high}(\alpha)(1-\epsilon/2),$$

when

$$\begin{aligned}
\epsilon &\in (0,2+I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})),\\
\delta &< \delta_2.
\end{aligned}$$

Note that for $\epsilon\geq 2+I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})$, $I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z})-\epsilon<0$, $1-\epsilon<0$, $\epsilon>1$, and there is nothing to prove.

In step 3, the wiretapper's decoding process is considered. In the wiretapper's decoding process, the wiretapper must select a parameter ξ for the joint typical decoder. If $\xi\neq\delta$, the set

$T_U^N(\delta) \neq T_U^N(\xi)$, and the decoder is not reliable. Hence, the wiretapper should select $\xi = \delta$, and that is what we will assume in the following analysis.

Firstly, the entropy of the codeword conditioned on bin j and the wiretapper's observation is related to the probability of error P_B in the wiretapper's bin-decoding process through random variable χ defined as

$$\chi = \begin{cases} 1 & \text{if } \psi(Z^N) \neq U_c \\ 0 & \text{if } \psi(Z^N) = U_c \end{cases},$$

where $\psi(Z^N)$ is the wiretapper's decoding function. Then, the probability of error in the wiretapper's bin-decoding process is bounded using the union bound, resulting in constraints on δ and N.

Now, the entropy of the codeword conditioned on bin j and the wiretapper's observation Z^N is related to the probability of error in the bin-decoding process.

$$\begin{aligned}
H(U_c|S^K, Z^N) &= H(\chi, U_c|S^K, Z^N) - H(\chi|U_c, S^K, Z^N) \\
&= H(\chi|S^K, Z^N) + H(U_c|\chi, S^K, Z^N) - H(\chi|U_c, S^K, Z^N) \\
&\overset{(12)}{\leq} H(\chi|S^K) + H(U_c|\chi, S^K, Z^N) \\
&= \sum_{j=1}^{2^{NR_{high}(\alpha)}} \Pr\{S^K = j\}[H(\chi|S^K = j) \\
&\quad + p(\chi = 0|S^K = j)H(U_c|Z^N, S^K = j, \chi = 0) \\
&\quad + p(\chi = 1|S^K = j)H(U_c|Z^N, S^K = j, \chi = 1)] \\
&\leq \sum_{j=1}^{2^{NR_{high}(\alpha)}} \Pr\{S^K = j\}[h(P_B) + (1 - P_B)H(U_c|Z^N, S^K = j, \chi = 0) \\
&\quad + P_B H(U_c|Z^N, S^K = j, \chi = 1)] \\
&\leq \sum_{j=1}^{2^{NR_{high}(\alpha)}} \Pr\{S^K = j\}[h(P_B) + P_B \log(2^{N[I(\tilde{U};\tilde{Z}) - \epsilon_{UZ}]} - 1)] \\
&= h(P_B) + P_B \log(2^{N[I(\tilde{U};\tilde{Z}) - \epsilon_{UZ}]} - 1) \\
&\leq h(P_B) + P_B N[I(\tilde{U};\tilde{Z}) - \epsilon_{UZ}] \\
&\leq h(P_B) + P_B N I(\tilde{U};\tilde{Z}) \\
\frac{H(X_c|S^K, Z^N)}{R_{high}(\alpha)N} &\leq \frac{h(P_B) + P_B N I(\tilde{U};\tilde{Z})}{R_{high}(\alpha)N} \\
&\leq \epsilon/2,
\end{aligned}$$

for $N \geq L_3 = 2h(P_B)/\{R_{high}(\alpha)\epsilon - 2P_B I(\tilde{U};\tilde{Z})\}$, and $P_B < R_{high}(\alpha)\epsilon/2I(\tilde{U};\tilde{Z})$, where $h(P_B)$ is the binary entropy function. (12) follows from the fact that $H(\chi|S^K, Z^N) \leq H(\chi|S^K)$ and $H(\chi|U_c, S^K, Z^N) = 0$.

To bound P_B, define the following events associated with the wiretapper's bin-decoding process:

- $\mathcal{E}^{Z1}(j)$: in the wiretapper's bin-decoding process, there is no sequence u_c in bin j that is jointly typical the received sequence z^N.

- $\mathcal{E}^{Z2}(j)$: in the wiretapper's bin-decoding process, a sequence $u_c \neq u_c(j,k^*)$ is jointly typical with the received sequence z^N for u_c in bin j.

An error in wiretapper's bin-decoding process occurs when at least one of the events $\mathcal{E}^{Z1}(j)$ and $\mathcal{E}^{Z2}(j)$ occurs when there is no encoding error. Then, the probability of wiretapper's decoding error given the secret message is

$$\begin{aligned} P_B &= p(\chi = 1|S^K = j) \\ &\leq \Pr\{\mathcal{E}^{Z1}(j)|\mathcal{E}^X(j)^C\mathcal{E}^V(j)^C\} + \Pr\{\mathcal{E}^{Z2}(j)|\mathcal{E}^X(j)^C\mathcal{E}^V(j)^C\} \\ &= \Pr\{\mathcal{E}^{Z1}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\} + \Pr\{\mathcal{E}^{Z2}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}. \end{aligned}$$

The two probabilities $\Pr\{\mathcal{E}^{Z1}(j)|\mathcal{E}^X(j)^C\mathcal{E}^V(1)^C\}$ and $\Pr\{\mathcal{E}^{Z2}(j)|\mathcal{E}^X(j)^C\mathcal{E}^V(1)^C\}$ are evaluated as follows:

$\Pr\{\mathcal{E}^{Z1}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}$. By Lemma 2.5 with $\epsilon_1 = R_{high}(\alpha)\epsilon/6I(\tilde{U};\tilde{Z})$, there exists L_4 such that $\Pr\{(U_c, Z^N) \in T^N_{U,Z}(\delta)\} > 1 - R_{high}(\alpha)\epsilon/6I(\tilde{U};\tilde{Z})$ for a given $\delta > 0$ when $N \geq L_4$, which implies that $\Pr\{\mathcal{E}^{Z1}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\} \leq R_{high}(\alpha)\epsilon/6I(\tilde{U};\tilde{Z})$ for a given $\delta > 0$ and $N \geq L_4$.

$\Pr\{\mathcal{E}^{Z2}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}$. By the same analysis as in the evaluation of the probability $\Pr\{\mathcal{E}^{Z2}(1,1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\}$ in the proof of Theorem 4.2 with a replacement of $R_{cam}(\alpha)$ by $R_{high}(\alpha)$, $\Pr\{\mathcal{E}^{Z2}(1)|\mathcal{E}^X(1)^C\mathcal{E}^V(1)^C\} \leq \epsilon/3$ for $0 < \delta < \min\{\delta_3, \epsilon_{UZ}/3\}$ and $N \geq L_5$.

For the coding strategy to work, given an $\epsilon \in (0, 2+I(\tilde{U};\tilde{Y})-I(\tilde{U};\tilde{Z}))$, pick $0 < \epsilon_{UZ} < \Delta I(1)$, and δ such that $0 < \delta < \min\{1, \delta_0, \delta_1, \delta_2, \delta_3, \epsilon_{UZ}/3\}$, and pick $N \geq N_{min} = \max\{L_0, L_1, L_2, L_3, L_4, L_5\}$.

Finally, we must ensure that $\Delta I(\alpha_{high}) > 0$, where $\Delta I(\alpha_{high})$

$$\begin{aligned} &= \frac{1}{2}\log\left[\frac{P(P+Q+N_1+N_2)}{(P+\alpha^2_{high}Q)(P+Q+N_1+N_2)-(P+\alpha_{high}Q)^2}\right] \\ &= \frac{1}{2}\log\left[\frac{P(P+Q+N_1+N_2)}{(P+Q)(P+Q+N_1+N_2)-(P+Q)^2}\right] \end{aligned}$$

$\Delta I(\alpha_{high}) > 0$ when

$$\begin{aligned} P(P+Q+N_1+N_2) &> (P+Q)(P+Q+N_1+N_2)-(P+Q)^2 \\ 0 &> Q(P+Q+N_1+N_2)-P^2-2PQ-Q^2 \\ &< P^2+PQ-Q(N_1+N_2) \end{aligned}$$

which is true when

$$P > -\frac{Q}{2}+\frac{\sqrt{Q^2+4Q(N_1+N_2)}}{2}.$$

■

4.4 Performance Comparisons

4.4.1 Behavior of the Leakage Function

Given the system parameters P, Q, N_1 and N_2, the leakage function associated with the parameters can be derived. Based on the leakage function, the parameter α can be selected appropriately to optimize the rate of transmission while maintaining the perfect secrecy. Below, the behavior of the leakage function is examined in details by looking at some of its properties.

Definition of the leakage function for GWCSI.

$$\begin{aligned}\Delta I(\alpha) &= I(\tilde{U};\tilde{Z}) - I(\tilde{U};V) \\ &= \frac{1}{2}\log\left[\frac{P(P+Q+N_1+N_2)}{(P+\alpha^2 Q)(P+Q+N_1+N_2)-(P+\alpha Q)^2}\right].\end{aligned}$$

The leakage function is positive when $\alpha = 0$

$$\begin{aligned}\Delta I(0) &= \frac{1}{2}\log\left[\frac{P(P+Q+N_1+N_2)}{P(P+Q+N_1+N_2)-P^2}\right] \\ &= \frac{1}{2}\log\left[\frac{P+Q+N_1+N_2}{Q+N_1+N_2}\right] > 0\end{aligned}$$

The value of the leakage function is zero at $\alpha = \alpha_0$, where

$$\alpha_0 = \frac{P}{P+N_1+N_2}\left[1+\sqrt{\frac{P+Q+N_1+N_2}{Q}}\right].$$

The leakage function has its maximum value at $\alpha = \alpha_w$.

$$\begin{aligned}\frac{d\Delta I(\alpha)}{d\alpha} &= \frac{1}{2\ln(2)}\left[\frac{(P+\alpha^2 Q)(P+Q+N_1+N_2)-(P+\alpha Q)^2}{P(P+Q+N_1+N_2)}\right]\times \\ &\quad -\left[\frac{P(P+Q+N_1+N_2)\{2\alpha Q(P+Q+N_1+N_2)-2Q(P+\alpha Q)\}}{\{(P+\alpha^2 Q)(P+Q+N_1+N_2)-(P+\alpha Q)^2\}^2}\right] \\ &= -\frac{1}{\ln(2)}\left[\frac{\alpha Q(P+Q+N_1+N_2)-Q(P+\alpha Q)}{(P+\alpha^2 Q)(P+Q+N_1+N_2)-(P+\alpha Q)^2}\right].\end{aligned}$$

Setting $d\Delta I(\alpha)/d\alpha = 0$,

$$\begin{aligned}\alpha(P+Q+N_1+N_2) &= P+\alpha Q \\ \alpha &= \alpha_w = \frac{P}{P+N_1+N_2} < \alpha_0.\end{aligned}$$

The leakage function has a positive value at $\alpha = 0$ and increases to its maximum value at $\alpha = \alpha_w$ then decreases to 0 at $\alpha = \alpha_0$. The leakage of the secret message to the wiretapper is therefore maximized if $\alpha = \alpha_w$. Note here that when $\alpha = \alpha_w$, the capacity of the combined main-wiretap channel is reached and is equal to $\frac{1}{2}\log[\frac{P+N_1+N_2}{N_1+N_2}]$ while the secret message rate in the main channel is below the capacity of the main channel ($\frac{1}{2}\log[\frac{P+N_1}{N_1}]$), which is achieved when $\alpha = \alpha_{cam}$.

There are three types of the leakage function characterized by the parameter α_0. The first type of the leakage function has $0 < \alpha_0 \leq \alpha_{cam}$. In this situation, the leakage function becomes non-positive when $\alpha \geq \alpha_0$ as seen in Figure 4.7. The second type of the leakage function has $\alpha_{cam} < \alpha_0 \leq 1$; the values of the function are positive at $\alpha = \alpha_{cam}$ and non-positive at $\alpha = 1$, respectively. See Figure 4.8. The third type has $\alpha_0 > 1$, with a positive value of the function at $\alpha = 1$. See Figure 4.9. This characterization will be useful for selecting mode of operation that is appropriate for a given set of the parameters as described in the next section.

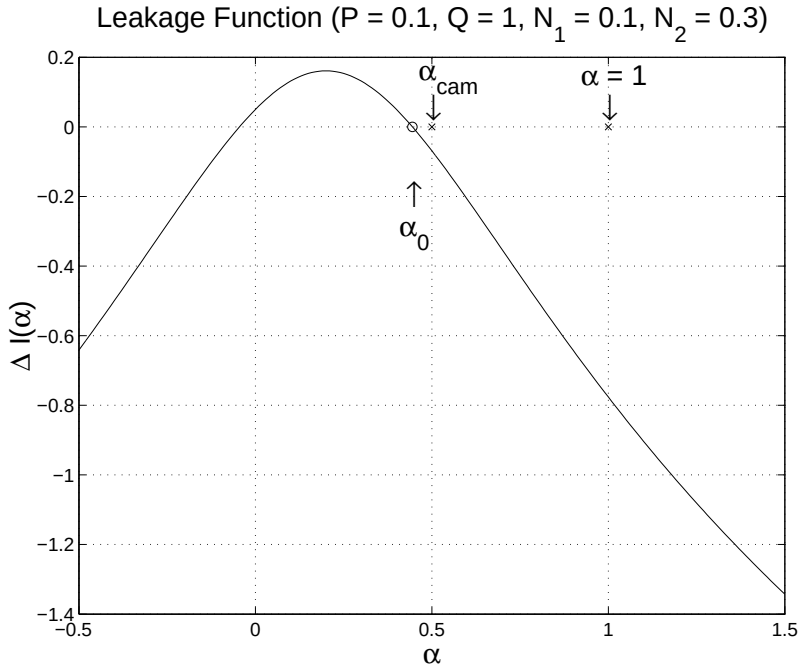

Figure 4.7: A leakage curve for the optimal camouflage mode.

4.4.2 A Characterization of GWCSI

We define P_{cam} and P_{high} for the characterization of GWCSI as

$$\begin{aligned} P_{cam} &= -N_1 - \frac{Q}{2} + \frac{\sqrt{Q^2 + 4QN_2}}{2} \\ P_{high} &= -\frac{Q}{2} + \frac{\sqrt{Q^2 + 4Q(N_1 + N_2)}}{2}. \end{aligned}$$

The characterization of the channel can be done based on the analysis of the leakage function and the proposed modes of operation, given a set of parameters P, Q, N_1 and N_2, into:

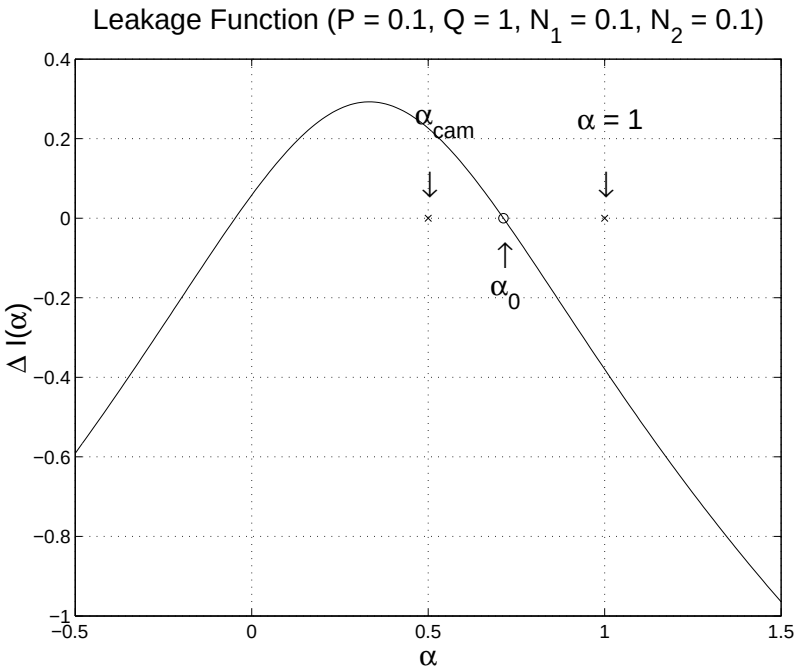

Figure 4.8: A leakage curve for the efficient camouflage mode.

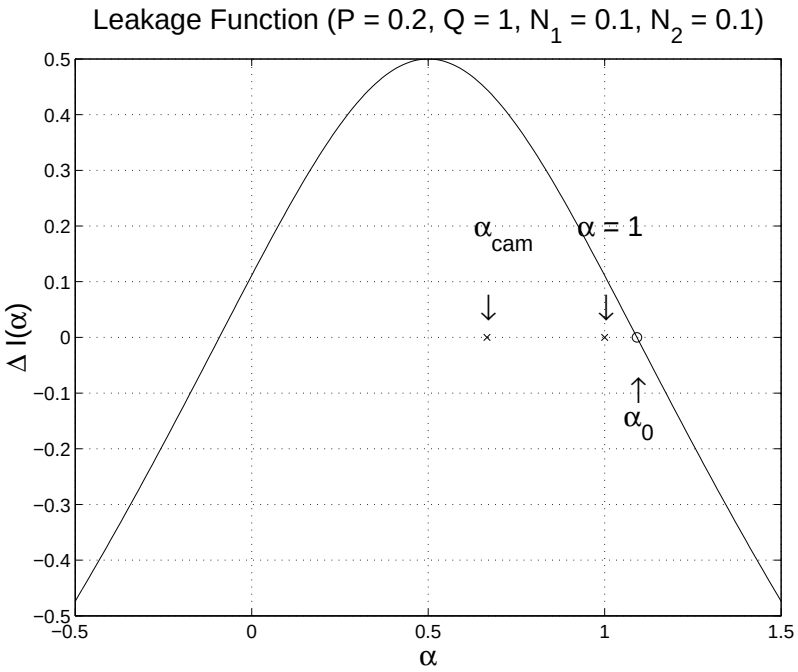

Figure 4.9: A leakage curve for the high-power mode.

1. The channel that is suitable for the optimal camouflage mode. The parameters associated with the channel of this type must satisfy the condition: $0 < P \leq P_{cam}$.

2. The channel that is suitable for the high-power mode. The parameters associated with the channel of this type must satisfy the condition: $P > P_{high}$.

3. The channel that is suitable for the efficient camouflage mode. The parameters associated with the channel of this type must satisfy the condition: $P_{cam} < P \leq P_{high}$

Hence, a mode of operation together with an appropriate α can be selected by comparing the given parameter P to P_{cam} and P_{high} calculated from Q, N_1 and N_2.

If the system parameters are adjustable, the channel can be tuned so that the desired mode of operation can be employed. For instance, if P is adjustable and $P_{cam} > 0$, all of the three modes of operation can be used based on the selection of P; thereby allowing the tradeoff between the average power constraint and the rate of transmission while maintaining perfect secrecy as illustrated in Figure 4.10. The rate corresponding to $P \leq P_{cam}$ is $R_{cam}(\alpha_{cam})$ in the optimal camouflage mode. The rate corresponding to $P_{cam} \leq P \leq P_{high}$ is $R_{cam}(\alpha_0)$ in the efficient camouflage mode. The rate corresponding to $P > P_{high}$ is $R_{high}(\alpha_{high})$ in the high-power mode.

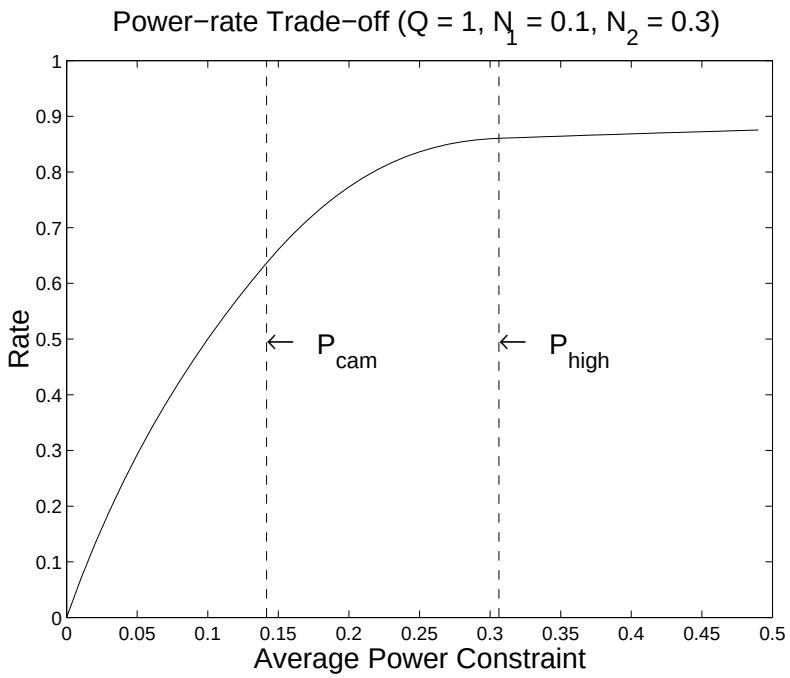

Figure 4.10: The power-rate tradeoff curve when $P_{cam} > 0$ at perfect secrecy.

On the other hand, if P is adjustable and $P_{cam} \leq 0$, either the efficient mode or the high power mode can be selected based on the average power constraint P, and the tradeoff between the average power constraint and the rate of transmission while maintaining perfect secrecy is illustrated in Figure 4.11.

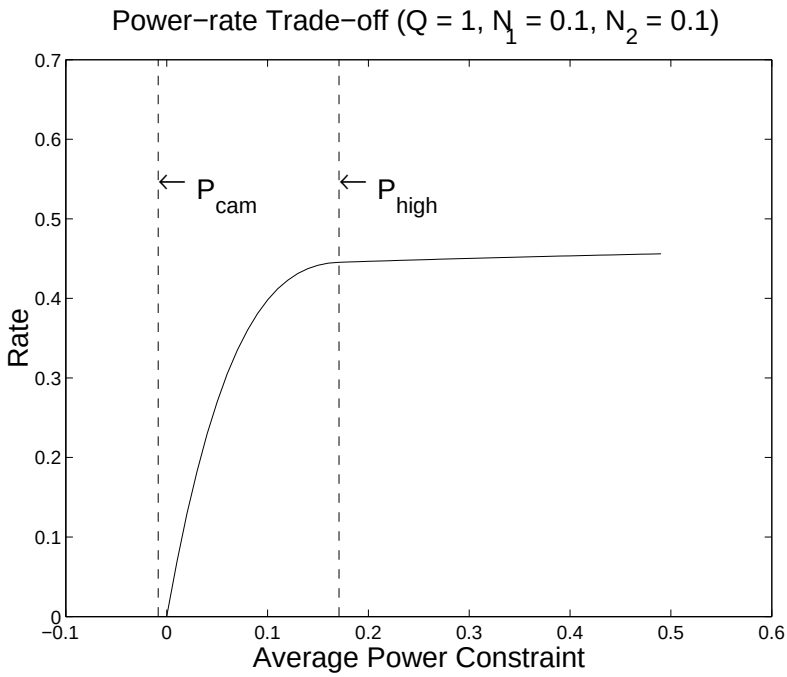

Figure 4.11: The power-rate tradeoff curve when $P_{cam} \leq 0$ at perfect secrecy.

The curves in Figure 4.10 and 4.11 are continuous and can be described as follows. The power-rate tradeoff relation at perfect secrecy is

$$\text{Rate} \quad = \quad \begin{cases} \frac{1}{2}\log\left[\frac{P+N_1}{N_1}\right] & \text{for } 0 \leq P \leq P_{cam} \text{ and } \alpha = \alpha_{cam} \\ \frac{1}{2}\log\left[\frac{P(P+Q+N_1)}{(P+\alpha_0^2 Q)(P+Q+N_1)-(P+\alpha_0 Q)^2}\right] & \text{for } P_{cam} \leq P \leq P_{high} \text{ and } \alpha = \alpha_0 \\ \frac{1}{2}\log\left[\frac{(P+Q+N_1)(N_1+N_2)}{(P+Q+N_1+N_2)N_1}\right] & \text{for } P_{high} \leq P \text{ and } \alpha = \alpha_{high} \end{cases} ,$$

The curve is continuous since at $P = P_{cam}$, $\alpha_0 = \alpha_{cam}$ while $\alpha_0 = \alpha_{high}$ at $P = P_{high}$.

4.5 An Achievable Region for GWCSI

Based on the camouflage and the high-power modes and the time-sharing lemma (Lemma 4.13), we prove an achievable region for the GWCSI. The achievability of a rate-equivocation pair is defined in the same way as in the case of GWC. Before proving the achievable region, we will need the following lemma.

Lemma 4.14 *For the Gaussian wiretap channel with side informatin with the parameters*

(P, Q, N_1, N_2), the pair (C_M, d_C) is achievable, where

$$\begin{aligned} C_M &= \frac{1}{2}\log\left[\frac{P+N_1}{N_1}\right] \\ d_C &= 1 - \frac{1}{2C_M}\log\left[\frac{(P(P+N_1)^2+P^2Q)(P+Q+N_1+N_2)}{(P(P+N_1)^2+P^2Q)(P+Q+N_1+N_2)-P^2(P+Q+N_1)^2}\right]. \end{aligned}$$

Proof: The main channel is the dirty-paper channel; hence the rate C_M is the capacity of the main channel and is achievable for arbitrarily small probability of error by the coding strategy proposed by Costa in [3]. Only the equivocation is left to be proved. We will use the same notation as that in the description of the camouflage mode and show that, at $\alpha = \alpha_{cam}$,

1. $H(S^K|Z^N) \geq N[I(U;Y) - I(\tilde{U};V) - I(U;Z) - \epsilon_{UV}]$;

2. $I(U;Y) - I(\tilde{U};V) - I(U;Z) - \epsilon_{UV} \geq (C_M - \epsilon)\left[1 - \frac{I(\tilde{U};\tilde{Z})}{I(\tilde{U};\tilde{Y})-I(\tilde{U};V)} - \epsilon\right]$.

Combining the two steps,

$$\begin{aligned} \frac{H(S^K|Z^N)}{H(S^K)} &\geq \frac{1}{(C_M-\epsilon)N} N[I(U;Y) - I(\tilde{U};V) - I(U;Z) - \epsilon_{UV}] \\ &\geq \frac{1}{(C_M-\epsilon)N}(C_M-\epsilon)N\left[1 - \frac{I(\tilde{U};\tilde{Z})}{I(\tilde{U};\tilde{Y}) - I(\tilde{U};V)} - \epsilon\right] \\ &= 1 - \frac{I(\tilde{U};\tilde{Z})}{C_M} - \epsilon \\ &= d_C - \epsilon \end{aligned}$$

since

$$\begin{aligned} d_C &= 1 - \frac{I(\tilde{U};\tilde{Z})}{C_M} \\ &= 1 - \frac{1}{2C_M}\log\left[\frac{((P+N_1)^2+PQ)(P+Q+N_1+N_2)}{((P+N_1)^2+PQ)(P+Q+N_1+N_2)-P(P+Q+N_1)^2}\right] \\ C_M &= I(\tilde{U};\tilde{Y}) - I(\tilde{U};V) \\ &= \frac{1}{2}\log\left[\frac{P}{P+N_1}\right] \end{aligned}$$

Proceeding to the first step,

$$\begin{aligned}
H(S^K|Z^N) &= H(S^K, Z^N) - H(Z^N) \\
&= H(S^K, U_c, Z^N) - H(U_c|S^K, Z^N) - H(Z^N) \\
&= H(S^K|U_c, Z^N) + H(U_c|Z^N) - H(U_c|S^K, Z^N) \\
&\overset{(1)}{=} H(U_c|Z^N) - H(U_c|S^K, Z^N) \\
&\overset{(2)}{\geq} H(U_c|Z^N) - N[I(\tilde{U};V) + \epsilon_{UV}] \\
&\overset{(3)}{\geq} H(U_c|Z^N) - H(U_c|Y^N) - N[I(\tilde{U};V) + \epsilon_{UV}] \\
&= [H(U_c, Z^N) - H(Z^N)] - [H(U_c, Y^N) - H(Y^N)] - N[I(\tilde{U};V) + \epsilon_{UV}] \\
&= [H(U_c) + H(Z^N|U_c) - H(Z^N)] - [H(U_c) + H(Y^N|U_c) - H(Y^N)] \\
&\quad -N[I(\tilde{U};V) + \epsilon_{UV}] \\
&\overset{(4)}{=} [H(Z^N|U_c) - H(Z^N|U^N) - H(U^N) + H(U^N|Z^N)] \\
&\quad -[H(Y^N|U_c) - H(Y^N|U^N) - H(U^N) + H(U^N|Y^N) - N[I(\tilde{U};V) + \epsilon_{UV}] \\
&\overset{(5)}{=} H(U^N|Z^N) - H(U^N|Y^N) - N[I(\tilde{U};V) + \epsilon_{UV}] \\
&= I(U^N;Y^N) - I(U^N;Z^N) - N[I(\tilde{U};V) + \epsilon_{UV}] \\
&= N[I(U;Y) - I(\tilde{U};V) - I(U;Z) - \epsilon_{UV}].
\end{aligned}$$

(1) follows from the fact that $H(S^K|U_c, Z^N) = 0$. (2) follows from the fact that there are $2^{N[I(\tilde{U};V)+\epsilon_{UV}]}$ auxiliary codewords in a bin. (3) follows from the fact that $H(U_c|Y^N) \geq 0$. (4) follows from the fact that $H(Z^N) = H(U^N) + H(Z^N|U^N) - H(U^N|Z^N)$ and $H(Y^N) = H(U^N) + H(Y^N|U^N) - H(U^N|Y^N)$. (5) follows from the fact that $H(Y^N|U_c) = H(Y^N|U^N)$ and $H(Z^N|U_c) = H(Z^N|U^N)$.

Proceeding to the second step, define $g_{LHS}(\delta) = I(U;Y) - I(\tilde{U};V) - I(U;Z) + I(\tilde{U};\tilde{Z}) - \epsilon_{UV}$ and $g_{RHS}(\delta) = C_M(1-\epsilon) = [I(\tilde{U};\tilde{Y}) - I(\tilde{U};V)](1-\epsilon)$. Imposing the condition $I(\tilde{U};\tilde{Y}) - I(\tilde{U};V) - \epsilon_{UV} > C_M(1-\epsilon)$, then

$$\begin{aligned}
g_{LHS}(0) &= I(\tilde{U};\tilde{Y}) - I(\tilde{U};V) - \epsilon_{UV} \\
g_{RHS}(\delta) &= C_M(1-\epsilon) < g_{LHS}(0).
\end{aligned}$$

Hence, $g_{LHS}(\delta)$ is continuous in δ with the value of $I(\tilde{U};\tilde{Y}) - I(\tilde{U};V) - \epsilon_{UV}$ at $\delta = 0$ while $g_{RHS}(\delta)$ is continuous and constant in δ with the value less than $I(\tilde{U};\tilde{Y}) - I(\tilde{U};V) - \epsilon_{UV}$. If the two curves do not intersect, then there is no constraint on δ. However, if the two curves intersect, then we have the constraint $0 < \delta < \delta^*$, where δ^* is the smallest value such that the two curves intersect. Then the condition on δ becomes $0 < \delta < \delta_0$, where $\delta_0 = \min\{\delta^*, \infty\}$. Consequently,

$$I(U;Y) - I(\tilde{U};V) - I(U;Z) + I(\tilde{U};\tilde{Z}) - \epsilon_{UV} \quad \geq \quad [I(\tilde{U};\tilde{Y}) - I(\tilde{U};V)][1-\epsilon],$$

and

$$\begin{aligned}
I(U;Y) - I(\tilde{U};V) - I(U;Z) - \epsilon_{UV} & \\
&\geq [I(\tilde{U};\tilde{Y}) - I(\tilde{U};V)][1-\epsilon] - I(\tilde{U};\tilde{Z}) \\
&= [I(\tilde{U};\tilde{Y}) - I(\tilde{U};V)]\left[1 - \frac{I(\tilde{U};\tilde{Z})}{I(\tilde{U};\tilde{Y}) - I(\tilde{U};V)} - \epsilon\right] \\
&> [I(\tilde{U};\tilde{Y}) - I(\tilde{U};V) - \epsilon]\left[1 - \frac{I(\tilde{U};\tilde{Z})}{I(\tilde{U};\tilde{Y}) - I(\tilde{U};V)} - \epsilon\right] \\
&= (C_M - \epsilon)\left[1 - \frac{I(\tilde{U};\tilde{Z})}{I(\tilde{U};\tilde{Y}) - I(\tilde{U};V)} - \epsilon\right]
\end{aligned}$$

when

$$\begin{aligned}
\epsilon_{UV} &< I(\tilde{U},\tilde{Y}) - I(\tilde{U},V) - C_M(1-\epsilon) = \epsilon C_M \\
\delta &< \delta_0.
\end{aligned}$$

The theorem follows by setting appropriate constants ϵ_{UV} and δ. ■

Theorem 4.4 *For the Gaussian wiretap channel with side information, a rate-equivocation pair (R,d) is achievable if*

$$\begin{aligned}
R &\leq C_M \\
d &\leq 1 \\
Rd &\leq \begin{cases} C_M & 0 < P \leq P_{cam} \\ \min\{C_M d_C, R_1\} & P_{cam} < P \leq P_{high} \\ \min\{C_M d_C, R_2\} & P > P_{high} \end{cases}
\end{aligned}$$

where

$$\begin{aligned}
C_M &= \frac{1}{2}\log\left[\frac{P}{P+N_1}\right] \\
d_C &= 1 - \frac{1}{2C_M}\log\left[\frac{(P(P+N_1)^2 + P^2Q)(P+Q+N_1+N_2)}{(P(P+N_1)^2 + P^2Q)(P+Q+N_1+N_2) - P^2(P+Q+N_1)^2}\right] \\
R_1 &= \frac{1}{2}\log\left[\frac{P(P+Q+N_1)}{(P+\alpha_0^2Q)(P+Q+N_1) - (P+\alpha_0 Q)^2}\right] \\
R_2 &= \frac{1}{2}\log\left[\frac{(P+Q+N_1)(N_1+N_2)}{(P+Q+N_1+N_2)N_1}\right] \\
\alpha_0 &= \frac{P}{P+Q+N_1}\left[1 + \sqrt{\frac{P+Q+N_1+N_2}{Q}}\right].
\end{aligned}$$

Proof: For $0 < P \leq P_{cam}$, the optimal camouflage mode can be used to achieve the pair $(C_M, 1)$. Using codes that perform worse than that in the camouflage mode achieve the pairs dominated by $(C_M, 1)$.

For $P_{cam} < P \leq P_{high}$, if $R_1 \leq C_M d_C$, time-share the code in the efficient camouflage mode with the code achieving the pair $(C_M, R_1/C_M)$. Otherwise, time-share the codes achieving the pairs $(C_M d_c, 1)$ and (C_M, d_C).

For $P > P_{high}$, if $R_2 \leq C_M d_C$, time-share the code in the high-power mode with the code achieving the pair $(C_M, R_2/C_M)$. Otherwise, time-share the codes achieving the pairs $(C_M d_c, 1)$ and (C_M, d_C). ■

4.5.1 Bounds on the Rate at Perfect Secrecy

The rate of transmission at perfect secrecy depends only on the average power constraint parameter P when Q, N_1 and N_2 are given. An upper bound on the rate can be obtained when considering the following:

$$\begin{aligned}
H(S^K|Z^N) &\overset{(1)}{\leq} H(S^K|Z^N) + [H(S^K|Y^N) - H(S^K|Z^N, Y^N)] \\
&\overset{(2)}{\leq} H(S^K|Z^N) + [H(S^K|\hat{S}^K) - H(S^K|Z^N, Y^N)] \\
&\overset{(3)}{\leq} H(S^K|Z^N) + [f(P_e) - H(S^K|Z^N, Y^N)] \\
&= H(S^K|Z^N) - H(S^K|Z^N, Y^N) + f(P_e) = I(S^K; Y^N|Z^N) + f(P_e) \\
&\overset{(4)}{=} H(S^K|Z^N) - H(S^K|Y^N) + f(P_e) \\
&= I(S^K; Y^N) - I(S^K; Z^N) + f(P_e)
\end{aligned}$$

where

$$f(P_e) = -P_e \log(P_e) - (1 - P_e) \log(1 - P_e) + K P_e \log(|\mathcal{S}|).$$

(1) follows from the fact that $H(S^K|Y^N) - H(S^K|Z^N, Y^N) \geq 0$. (2) follows from the data processing inequality. (3) follows from Fano's inequality. (4) follows from the fact that Z^N is a degraded version of Y^N.

The simplest upper bound follows from the fact that $I(S^K; Z^N) \geq 0$.

$$\begin{aligned}
H(S^K|Z^N) &\leq I(S^K; Y^N) + f(P_e) \\
&\leq \frac{1}{2} \log \left[\frac{P + N_1}{N_1} \right] + f(P_e).
\end{aligned}$$

If there exists a code for a reliable communication at rate $H(S^K)/N$, $f(P_e)$ tends to 0 as N approaches ∞. This upper coincides with the capacity of the dirty-paper channel when the noise in the main channel has variance N_1.

A lower bound can be obtained by considering the interference as an additional noise in the main channel unknown to the sender. Accordingly, this is the case of Gaussian wiretap channel with $\sigma^2_{\eta_1} = Q + N_1$ as opposed to N_1. At perfect secrecy, the rate of transmission is

$$I(\tilde{X}; \tilde{Y}) - I(\tilde{X}; \tilde{Z}) = \frac{1}{2} \log \left[\frac{(P + Q + N_1)(Q + N_1 + N_2)}{(Q + N_1)(P + Q + N_1 + N_2)} \right].$$

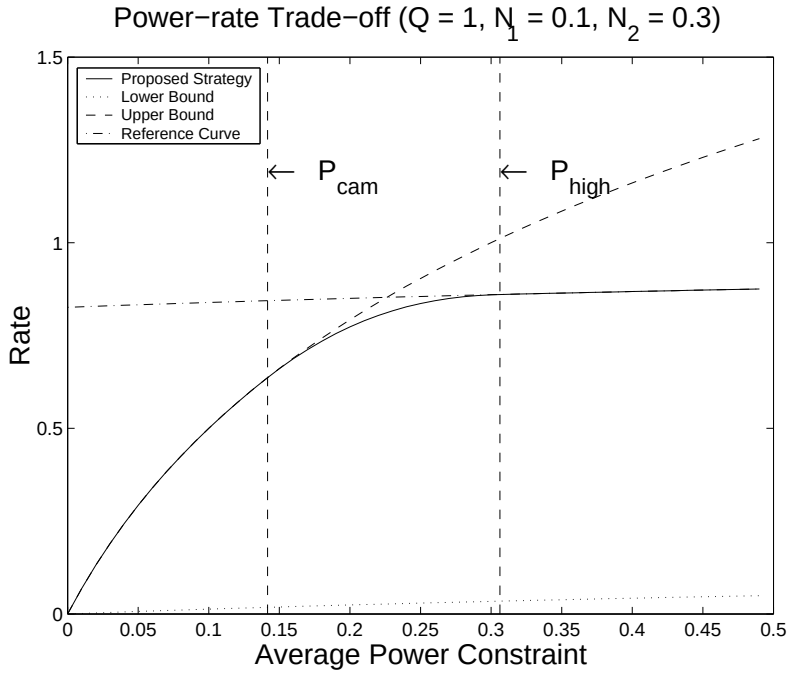

Figure 4.12: Bounds on power-rate tradeoff curve when $P_{cam} > 0$ at perfect secrecy.

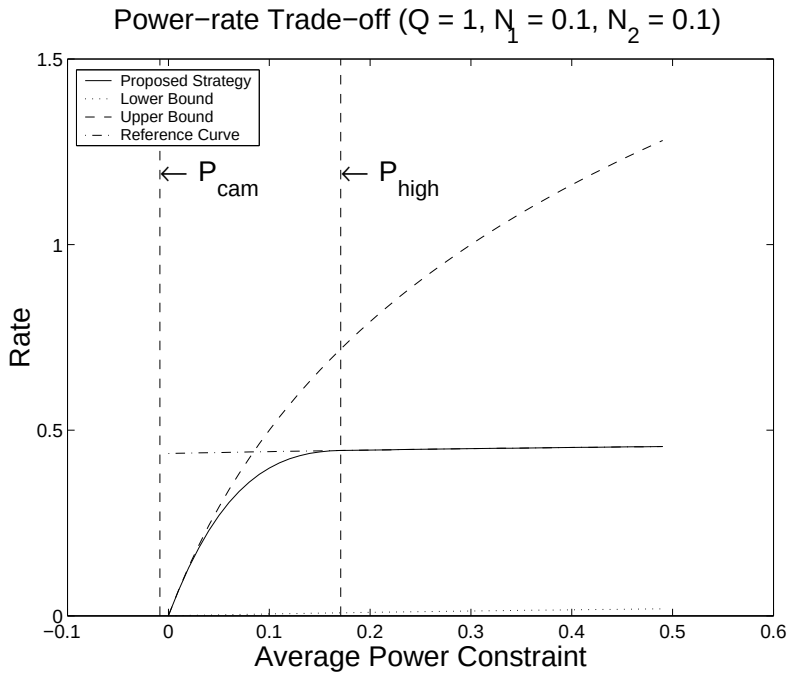

Figure 4.13: Bounds on power-rate tradeoff curve when $P_{cam} \leq 0$ at perfect secrecy.

As a reference, we consider the case of Gaussian wiretap channel with parameters $\sigma_X^2 \leq P+Q, \sigma_{\eta_1}^2 = N_1$ and $\sigma_{\eta_2}^2 = N_2$. The rate of transmission of this case is

$$I(\tilde{X};\tilde{Y}) - I(\tilde{X};\tilde{Z}) \quad = \quad \frac{1}{2}\log\left[\frac{(P+Q+N_1)(N_1+N_2)}{(N_1)(P+Q+N_1+N_2)}\right].$$

From Figures 4.12 and 4.13, we see that the curve of the proposed strategy coincides with the upper bound for $0 < P \leq P_{cam}$, and coincides with the reference curve when $P \geq P_{high}$, and is considerable higher than the lower bound.

4.6 Remarks

In this chapter, the interference-free mode based on the code-partitioning concept is proposed as a specific coding strategy for the Gaussian wiretap channel. It is extended to the camouflage and the high-power modes for the Gaussian wiretap channel with side information.

The introduction of leakage functions leads to a characterization of the GWCSI. The leakage functions can be used as a basis for selecting a mode of operation for a specific GWCSI. With this tool, a designer with flexible average power constraint can trade between the power and the rate of transmission using the camouflage mode and the high-power mode while maintaining asymptotic perfect secrecy.

The power-rate tradeoff at perfect secrecy curve is then compared to the upper and the lower bounds, which indicates that the system performs significantly better when the interference is known to the encoder in advance as opposed to treating it as noise in the channel.

Additionally, an achievable region for the GWCSI is derived based on time-sharing different codes, but the capacity region is still unknown.

Chapter 5

An Application of GWCSI

5.1 Introduction

The coding strategy for the Gaussian wiretap channel with side information finds it application in information hiding with the main concern in the secrecy of the information. The coding strategies proposed in the previous chapter can asymptotically achieve the perfect secrecy with tremendous complexity of the random coding technique. Although the random coding technique is rather impractical, it provides a limit of how far one can goes and, often enough, a clue of how a practical scheme may be implemented. In this chapter, an implementation of the coding strategy for the dirty-paper channel proposed by Eggers et al. in [8] for image watermarking is reviewed. The implementation is straightforwardly extended to augment the secrecy of the message against the wiretapper in the situation of the Gaussian wiretap channel with side information.

5.2 Information Embedding Based on Structured Codebooks

In [8], Eggers et al. introduced a blind watermarking scheme based on structured codebooks. The scheme is based on the random coding strategy for the dirty-paper channel discussed in the previous chapter and employs a lattice-structured codebook to reduce the complexity. The watermarking scheme aims to embed information in a medium, which can be regarded as the interference in the dirty-paper channel. The medium is known to the encoder in advance and can be used as side information in the encoding process. A case of binary-alphabet messages of the scheme is reviewed here.

Let S represent a message such that $S \in \{0, 1\}$ to be communicated over a dirty-paper channel with interference $V \sim \mathcal{N}(0, Q)$ and noise $\eta_1 \sim \mathcal{N}(0, N_1)$ under an average power constraint P. Let $\mathcal{U}_0$ and $\mathcal{U}_1$ be sets of codewords corresponding to the message "0" and "1," respectively.

The sets are constructed as uniform quantizers with cell width $\alpha\omega$ so that

$$\mathcal{U}_0 = \{u = i\alpha\omega : i \text{ are integers}\} \quad (5.1)$$

$$\mathcal{U}_1 = \{u = i\alpha\omega + \frac{\alpha\omega}{2} : i \text{ are integers}\}. \quad (5.2)$$

The parameters α and ω will be specified later. The entire codebook, $C = \mathcal{U}_0 \cup \mathcal{U}_1$, is given to

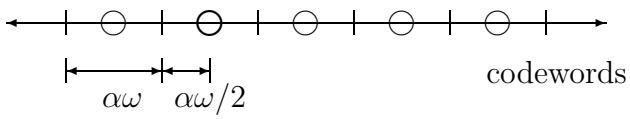

Figure 5.1: The codebook consists of the quantization points in $\mathcal{U}_0$ and $\mathcal{U}_1$.

both the sender and the recipient. In Figure 5.1, the points on the axis marked by the circles are the codewords in $\mathcal{U}_0$ representing the message "0," and the points of the axis marked by the vertical lines are the codewords in $\mathcal{U}_1$ representing the message "1."

Given an interference value v and a message s, the encoder uses the quantizer $\mathcal{U}_s/\alpha$ to quantize the interference so that the output of the quantizer and the quantizer error respectively are

$$\begin{aligned} v_q &= \mathcal{Q}(v, \frac{\mathcal{U}_s}{\alpha}) \\ e &= v_q - v. \end{aligned} \quad (5.3)$$

The encoder transmits αe. The recipient receives $y = \alpha e + v + \eta_1$ from the channel and decodes for the message by quantizing y with both $\mathcal{U}_0/\alpha$ and $\mathcal{U}_1/\alpha$ to obtain the corresponding quantization errors e_0 and e_1. The recipient declares "0" as the received message if $|e_0| \leq |e_1|$; otherwise, declares "1" as the received message.

$$e_0 = y - \mathcal{Q}(y, \frac{\mathcal{U}_0}{\alpha}) \quad (5.4)$$

$$e_1 = y - \mathcal{Q}(y, \frac{\mathcal{U}_1}{\alpha}) \quad (5.5)$$

The parameters ω and α are used to adjust the average input power and optimizing the transmission rate. The scaled quantization error is the input to the channel. It has been shown that, for a uniform quantizer, the mean squared error is $E[e^2] = \frac{(\alpha\omega)^2}{12}$. To achieve the average power constraint, let $\frac{(\alpha\omega)^2}{12} = P$. Given the interference variance Q and the noise variance N_1, the optimal ω that optimizes the rate is estimated numerically in [8] to be

$$\omega = \sqrt{12(P + 2.71N_1)}. \quad (5.6)$$

The corresponding α is

$$\alpha = \alpha_q = \sqrt{\frac{P}{12(P + 2.71N_1)}}. \quad (5.7)$$

The procedures for generating the codebook, encoding and decoding are summarized as follows:

1. *Generating the codebook.* Given an average power constraint P and noise variance N_1, calculate ω and α_q according to equations (5.6), and (5.7), respectively. Construct two uniform quantizers $\mathcal{U}_0$ and $\mathcal{U}_1$ according to equations (5.1) and (5.2). Distribute α_q to the sender and the two quantizers $\mathcal{U}_0/\alpha_q$ and $\mathcal{U}_1/\alpha_q$ to the sender and the recipient.

2. *Encoding.* To send a message s through a channel with interference v, calculate the quantization error of v with respect to $\mathcal{U}_s/\alpha_q$ according to equation (5.3). Transmit $\alpha_q e$.

3. *Decoding.* The recipient receives $y = \alpha_q e + v + \eta_1$ and calculates the quantization errors e_0 and e_1 of y with respect to $\mathcal{U}_0/\alpha_q$ and $\mathcal{U}_1/\alpha_q$, respectively. Declare "0" as the received message if $|e_0| \leq |e_1|$; otherwise, declares "1" as the received message.

This information embedding scheme is implemented as a basis for the extension to be described in the next section. The probability of error at the recipient depends on the Watermark-to-Noise Ratio (WNR) defined in [8] as $10\log_{10}(P/N_1)$. The performance of the system is simulated, and the result is similar to what was obtained in [8] as plotted in Figure 5.2.

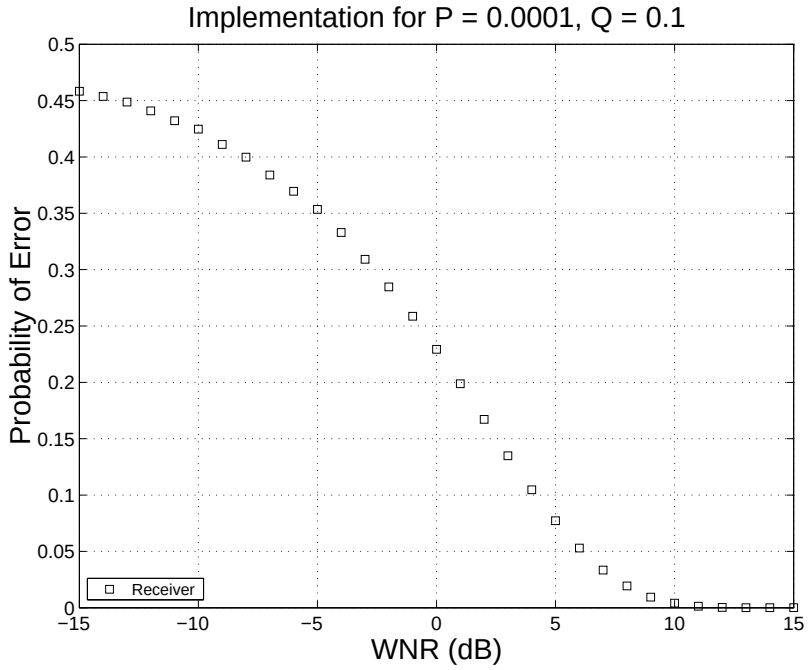

Figure 5.2: Performance of the information embedding scheme based on structured codebooks as a function of WNR.

An experiment of using the scheme with an image is performed on the 256-level gray-scaled image of Lena (Figure 5.3). The image is used as a medium for information embedding and is considered as the interference in the dirty-paper model. It consists of 256×256 pixels, and the variance of the pixel intensity is measured to be 0.0352. In a baseline experiment, 256×256 bits of information are embedded into the image, and the probability of error in detecting the information and the degradation of the image is observed. We fix the average power constraint P at 0.0001 and WNR at 12 dB, which implies that $N_1 = 6.3096 \times 10^{-6}$.

The resulting image at the recipient is shown in Figure 5.4. The recipient sees the image with Signal-to-Noise Ratio (SNR) 34.3296 dB and can detect embedded information with probability of error 4.1199×10^{-4}. In the next section, we will enhance the secrecy of the embedded information when there is an adversary observing the image being transmitted based on the GWCSI model.

The Original Image

Figure 5.3: The orignal image, with 256-level gray scale and 256×256 pixels, used in the experiments.

The Embeded Image (P = 0.0001, WNR = 12).

SNR = 34.3296 dB.

Figure 5.4: Information embedded image based on a structured codebook.

5.3 Information Embedding with Secrecy in Images

The information embedding scheme reviewed in the previous section can be extended based on the model of Gaussian wiretap channel with side information to enhance the secrecy of the information embedded. Two approaches of the extension are considered. The direct approach makes use of a key that is regarded as the noise in the wiretap channel while the indirect approach uses an interference mismatch as the noise in the wiretap channel.

5.3.1 Direct Approach

A direct extension of the information embedding scheme based on structured codebooks is done by introducing a key that is known to both the sender and the recipient. The key is perceived as an additional noise by the wiretapper. This can be straightforwardly implemented by adding the key to $\alpha_q e$ before the transmission; hence, what the recipient observes is $y = \alpha_q e + \eta_2 + v + \eta_1$, where η_2 is the key available to the sender and the recipient. Before the detection, the recipient subtracts the key from the received value and decodes for the message. If the variance of the key is high enough, the perfect secrecy of the message is asymptotically possible as shown in the case of the Gaussian wiretap channel with side information in the previous chapter. The procedure for binary message embedding using this extended scheme is described below.

1. *Generating the codebook.* Construct the two uniform quantizers $\mathcal{U}_0$ and $\mathcal{U}_1$ as before. Generate a key sequence η_2^N of the same length as that of the message i.i.d. according to $\mathcal{N}(0, N_2)$. Distribute α_q to the sender, the key sequence η_2^N and the two quantizers $\mathcal{U}_0/\alpha_q$ and $\mathcal{U}_1/\alpha_q$ to the sender and the recipient.

2. *Encoding.* To send a message s through a channel with interference v, calculate the quantization error of v with respect to $\mathcal{U}_s/\alpha_q$ according to equation (5.3). Transmit $\alpha_q e + \eta_2$.

3. *Decoding.* The recipient observes $y = \alpha_q e + \eta_2 + v + \eta_1$ and calculates the quantization errors e_0 and e_1 of $y - \eta_2$ with respect to $\mathcal{U}_0/\alpha_q$ and $\mathcal{U}_1/\alpha_q$, respectively. Declare "0" as the received message if $|e_0| \leq |e_1|$; otherwise, declares "1" as the received message.

Implementing the direct approach with WNR = 12 dB $P = 0.0001$ and $Q = 0.1$, the probability of error at the wiretapper is plotted as a function of the key power as multiples of the average power constraint on the codewords in Figure 5.5. Note that, for a fixed WNR, the probability of error tends to 0.5 as the key power increases.

For a fixed key power $N_2 = 3P$, the equivocation of the message at the wiretapper increases as the power of the noise in the main channel approaches 0 i.e., WNR approaches ∞. However, the increase of the equivocation is not monotonic as shown by a simulation result in Figure 5.6.

We conjecture that the dip in the probability of error results from the fact that the codebook parameters ω and α are optimized for the rate of transmission, not for the equivocation. As seen in the analysis of the equivocation in the Gaussian wiretap channel with side information case, the codebook parameters crucially influence the equivocation at the wiretapper.

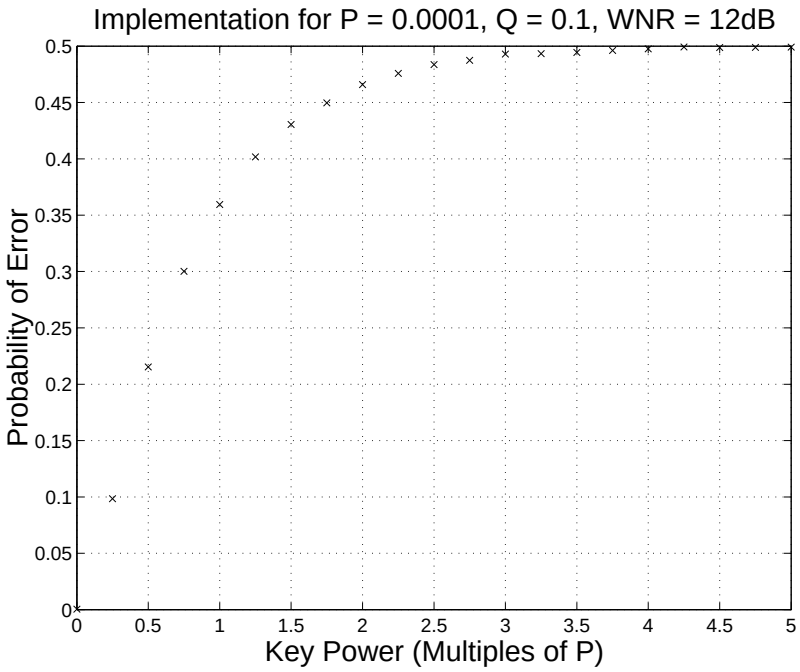

Figure 5.5: Performance of the secrecy-enhanced information embedding scheme based on structured codebooks as a function of key power at WNR = 12 dB.

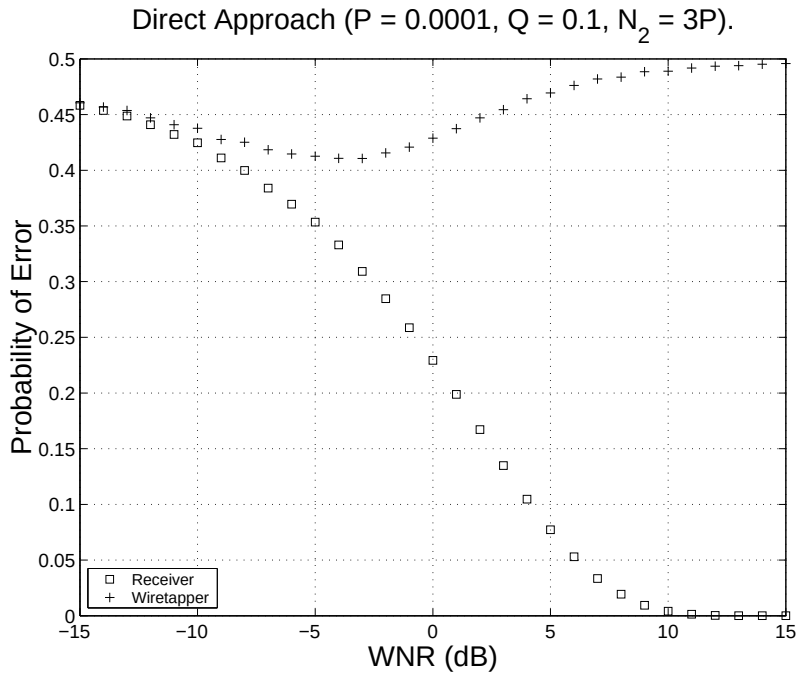

Figure 5.6: Performance of the direct approach secrecy-enhanced information embedding scheme based on structured codebooks as a function of WNR.

Applying the direct approach of the extension of the implementation to Lena image with a key sequence of generated i.i.d. from a normal distribution with mean zero and variance 0.0003, the recipient detects the embedded information with probability of error 3.2043×10^{-4} while the wiretapper detects the embedded information with probability of error 0.4870. The transmitted image however suffers a severe degradation (SNR = 28.1982 dB) due to the direct addition of the key since the output of the channel is $y = \alpha e + \eta_2 + v + \eta_1$. See Figure 5.7.

Direct Approach (P = 0.0001, N_2 = 0.0003, WNR = 12).

SNR = 28.1982 dB.

Figure 5.7: Direct approach secrecy-enhanced embedded image.

This straightforward extension increases the secrecy of the message at the cost of the degradation of the interference. In the next section, an indirect approach is proposed so that the quality of the image does not degrade as much as that in the direct approach.

5.3.2 Indirect Approach

An alternative extension integrates the key η_2 as part of the interference. The coding procedure is as described before except the encoding and the decoding processes since the interference does not affect the codebook generating process.

1. *Encoding.* To send a message s through a channel with interference v, calculate e, the quantization error of $v + \eta_2$ with respect to $\mathcal{U}_s/\alpha_q$, where $e = \mathcal{Q}(v + \eta_2, \frac{\mathcal{U}_s}{\alpha_q}) - v - \eta_2$. Transmit $\alpha_q e$.

2. *Decoding.* The recipient observes $y = \alpha e + v + \eta_1$ and calculates the quantization errors e_0 and e_1 of $y + \eta_2$ with respect to $\mathcal{U}_0/\alpha_q$ and $\mathcal{U}_1/\alpha_q$, respectively. Declare "0" as the received message if $|e_0| \leq |e_1|$; otherwise, declare "1" as the received message.

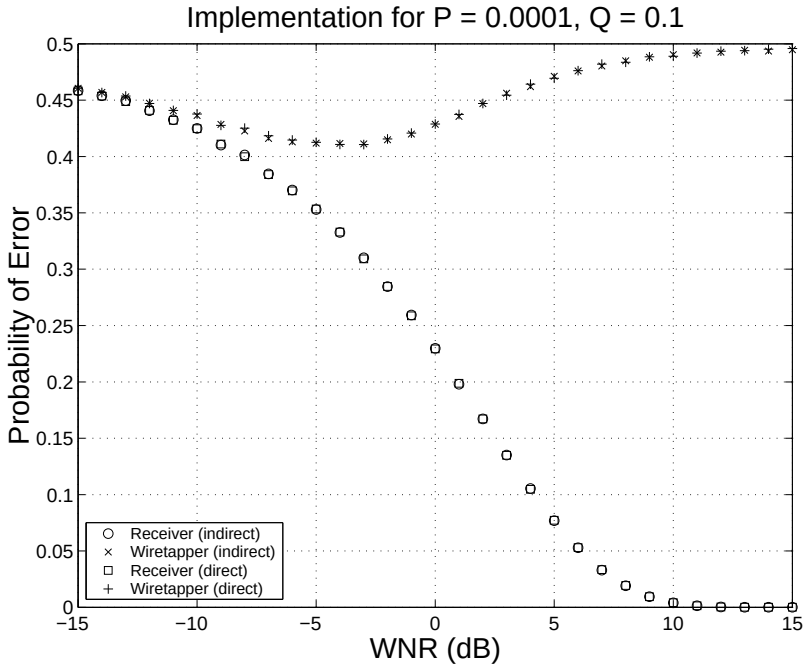

Figure 5.8: Comparison of the performances of the direct/indirect approach secrecy-enhanced information embedding scheme based on structured codebooks as a function of WNR.

Indirect Approach (P = 0.0001, N_2 = 0.0003, WNR = 12).

SNR = 34.1037 dB.

Figure 5.9: Indirect approach secrecy-enhanced embedded image.

We compare the performances by simulation in term of the probability of error as a function of WNR of the indirect approach and the direct approach in Figure 5.8, which indicates that there is virtually no difference in the performance of the two approaches.

Applying the indirect approach of the extension of the implementation to the Lena image with a key sequence generated i.i.d. from a normal distribution with mean zero and variance 0.0003, the recipient detects the embedded information with probability of error 0.0003 while the wiretapper detects the embedded information with probability of error 0.4906. The transmitted image suffers a small degradation (SNR = 34.1037 dB). See Figure 5.9.

5.4 Remarks

The Gaussian wiretap channel model is used for secret embedding in an image via extensions of a simple watermarking scheme based on the quantization technique. The direct extension adds the key sequence, available to only the sender and the recipient, directly to the image after the secret embedding process. The indirect extension regards the key sequence as another interference known to the encoder and uses it as side information in the encoding process to mitigate the degradation of the image.

The results of the experiments show that as the WNR increases, the equivocation of the embedded information improves while the power of the key sequence is fixed. On the contrary, the equivocation does not monotonically increase as the power of the key sequence increases. This situation may be improved by taking the equivocation into consideration when calculating the parameters α and ω.

Chapter 6

Wiretap Channel II with Side Information

6.1 Introduction

In this chapter, we revisit the wiretap channel II briefly introduced in Chapter 3 and relate the concept of Inverse/Dimension Length Profile (IDLP) to the coding strategy based on linear block codes used in [16] by Ozarow and Wyner to prove the achievability of the capacity region of the channel. This relationship provides a linkage between the performance of a system designed based on a generator matrix of a linear block code and a structure of the matrix, which can be used as a guideline for designing a good system for the channel based on a finite-length linear block code.

The wiretap channel II model is then considered under the hypothesis that the wiretapper has a capability to obtain a part of the uncoded message in addition to his capability to wiretap the main channel. The performance of the code designed for the wiretap channel II can be analyzed for the more capable wiretapper by employing the concept of Inverse Relative Dimension/Length Profile proposed by Luo et al. in [15] as an extension to the concept of IDLP.

Based on the concepts of IDLP and IRDLP, a guideline for designing good finite-length codes for the wiretap channel II and the wiretap channel II with side information are obtained.

6.2 A Coding Strategy for the Wiretap Channel II

Let S^K be the information to be kept secret, X^N be the output of the system encoder, $\tau^c \subseteq \{0, 1, 2, \ldots, N\}$ be an index set of size μ, and the wiretapped bits are $Z^\mu = X^N_{\tau^c}$, where S, X, and $Z \in \{0, 1\}$. The information source is assumed to be memoryless binary symmetric, that is $\Pr\{S = 0\} = \Pr\{S = 1\} = 1/2$. Let $\mathbf{A}$ be a $K \times N$ parity-check matrix known to the encoder, the decoder and the wiretapper. The coding strategy for the wiretap channel II as described

in Chapter 3 is called the coset coding method based on linear block codes and is summarized as follows:

1. *Generating the codebook.* Given a rank-K $K \times N$ binary parity-check matrix $\mathbf{A}$, solve the equation $\mathbf{A}\mathbf{x}^T = \mathbf{s}^T$ for each of the 2^K possible messages $\mathbf{s}^T$. Generate 2^K bins and uniquely associate each bin with a message. Put the 2^{N-K} solutions corresponding to each $\mathbf{s}^T$ into the bin associated with the message. Distribute the bins of length-N vectors as a codebook to the sender and the recipient.
2. *Encoding.* To send a message $\mathbf{s}^T$, randomly select a vector in the bin associated with the message and transmit the vector through the channel.
3. *Decoding.* The recipient observes the transmitted vector without error and declares the message associated with the bin containing the received vector as the received message.

The rate of this code is related to the dimension of the matrix $\mathbf{A}$ by $R = K/N$, and the probability of error is zero if there exists such a matrix. The equivocation of the message at the wiretapper can also be linked to the matrix $\mathbf{A}$. With the knowledge of $\mathbf{A}$ and $\mathbf{z}^\mu = \mathbf{x}_{\tau^c}$, the wiretapper can resolve the problem of determining the message into finding the set of solutions to the linear equation corresponding to each possible message $\mathbf{s}^T$:

$$\begin{aligned} \mathbf{A}_\tau \mathbf{x}_\tau^T + \mathbf{A}_{\tau^c}\mathbf{x}_{\tau^c}^T &= \mathbf{s}^T \\ \mathbf{A}_\tau \mathbf{x}_\tau^T &= \mathbf{s}^T + \mathbf{A}_{\tau^c}\mathbf{x}_{\tau^c}^T, \end{aligned} \tag{6.1}$$

where $\mathbf{A}_\tau$ denotes the submatrix of $\mathbf{A}$ consisting of columns of $\mathbf{A}$ indexed by τ; $\mathbf{x}_\tau$ denotes the subvector of $\mathbf{x}$ consisting of components of $\mathbf{x}$ indexed by τ, and $\tau^c = \{1, 2, \ldots, N\} \setminus \tau$. Equation (6.1) suggests that the number of possible messages $\mathbf{s}^T$ after observing $\mathbf{x}_{\tau^c}$ is 2^Λ, where $\Lambda = \mathrm{rank}(\mathbf{A}_\tau)$.

Furthermore, for each possible $\mathbf{s}^T$, there are $2^{N-\mu-\Lambda}$ valid solutions to the equation out of $2^{N-\mu}$ possible $\mathbf{x}_\tau^T$. Due to the coding strategy, all $\mathbf{x}^T$ are equally likely; hence, the probability that each message occurs is $2^{N-\mu-\Lambda}/2^{N-\mu} = 2^{-\Lambda}$, and the uncertainty of the message is $H(S^K|Z^\mu) = \Lambda$ for a given τ. Therefore, the best that the wiretapper can do is to select the subset τ^c such that Λ is minimized so that the equivocation is

$$\Delta = \min_{\tau^c : |\tau^c| = \mu} H(\mathbf{S}^K|\mathbf{Z}^\mu) = \min_{\tau : |\tau| = N-\mu} \mathrm{rank}(\mathbf{A}_\tau). \tag{6.2}$$

On the contrary, a system designer should seek to maximize the equivocation Δ by searching for a good matrix $\mathbf{A}$. To help the search, the concept of Inverse Dimension/Length Profile (IDLP) is brought in to link the equivocation of the message to the matrix as shown in the section 6.4. Next, we outline the important parts of the proof of the achievability of the capacity region of the wiretap channel II.

6.3 Capacity Region of the wiretap channel II

Ozarow and Wyner established the capacity region of the wiretap channel II in [16] to be the set of (R, α, d) such that $R \geq 0, \alpha \leq 1$, and

$$0 \leq d \leq \begin{cases} 1, & \text{for } 0 \leq \alpha \leq 1 - R, \\ \frac{1-\alpha}{R}, & \text{for } 1 - R \leq \alpha \leq 1. \end{cases} \tag{6.3}$$

The important parts of the proof of achievability are to show that the points $(R, 1-R, 1)$ is achievable for $0 \leq R \leq 1$ and $(R, \alpha, (1-\alpha)/R)$ is achievable for $0 \leq R \leq 1$ and $1-R \leq \alpha \leq 1$. These two parts are outlined here for they are crucial in understanding the analysis of the case, in which there is side information available to the wiretapper. The second part is the result of the following lemma.

Lemma 6.15 *([16]) Suppose that we are given an encoder/decoder pair with parameters N, K and P_e. Suppose that there are two wiretappers who have parameters $\mu = \mu_1, \mu_2$ and $\Delta = \Delta_1, \Delta_2$, respectively. If $\mu_2 \geq \mu_1$, then $\Delta_2 \geq \Delta_1 - (\mu_2 - \mu_1)$.*

Proof: Let $\tau_1^c \subseteq \tau_2^c \subseteq \{1, 2, \ldots, N\}$ where $|\tau_1^c| = \mu_1$, $|\tau_2^c| = \mu_2$. Let $Z_i = Z^{\mu_i} = X_{\tau_i^c}^N$ be the vector containing the components of X^N indexed by τ_i^c. Then,

$$\begin{aligned} H(S^K|Z_2) - H(S^K|Z_1) &= H(S^K|Z_2, Z_1) - H(S^K|Z_1) \\ &= -I(S^K; Z_2|Z_1) \\ &\geq -H(Z_2|Z_1) \geq -(\mu_2 - \mu_1), \end{aligned}$$

where the last inequality follows from $\tau_1^c \subseteq \tau_2^c$. Thus

$$\begin{aligned} H(S^K|Z_2) &\geq H(S^K|Z_1) - (\mu_2 - \mu_1) \\ &\geq \Delta_1 - (\mu_2 - \mu_1) \end{aligned} \tag{6.4}$$

from the definition of Δ. Minimizing (6.4) over all τ_2^c with $|\tau_2^c| = \mu_2$, yields

$$\Delta_2 \geq \Delta_1 - (\mu_2 - \mu_1).$$

■

Lemma 6.15 implies that

$$\frac{\Delta_2}{K} \geq \frac{\Delta_1}{K} - \frac{\mu_2/N - \mu_1/N}{K/N},$$

from which we conclude that (R, α_1, δ_1) is achievable implies that (R, α_2, δ_2) is achievable where $\alpha_2 \geq \alpha_1$ and

$$\delta_2 \geq \delta_1 - \left(\frac{\alpha_2 - \alpha_1}{R} \right).$$

The achievability proof of the triple $(R, 1-R, 1)$ for $0 \leq R \leq 1$ is based on the properties of large binary random matrices given by Lemma 6.16 and Lemma 6.17 below.

Lemma 6.16 *[16] Let $1 \leq K \leq N$, and let the $K \times N$ matrix $\mathbf{A}$ over $GF(2)$ be chosen at random with uniform distribution on the set of 2^{KN} binary $K \times N$ matrices. Then, for $1 \leq L \leq K$, $Pr\{rank(\mathbf{A}) < K - L\} \leq 2^{-(L+1)(N-K)+N}$.*

Proof: See *Lemma 5* in [16]. ■

Lemma 6.17 *[16] Let $1 \leq K \leq N$, and let $K \times N$ matrix $\mathbf{A}$ over GF(2) be chosen at random with uniform distribution on the set of 2^{KN} binary $K \times N$ matrices. Then*

$$Pr\{rank(\mathbf{A}) = K\} = \prod_{j=0}^{K-1}(1 - 2^{j-N}) \geq 1 - \frac{K2^{K-1-N}}{1 - 2^{K-1-N}}.$$

Proof: See *Lemma 6* in [16]. ■

It was shown that there exists a $K \times N$ matrix $\mathbf{A}$ such that, for $\epsilon > 0$ and sufficiently large N,

1. rank of any $K + \epsilon N$ columns of $\mathbf{A}$ is close to K, and
2. rank of $\mathbf{A}$ is K.

Hence, the matrix $\mathbf{A}$ can be used to generate a codebook for the wiretap channel II that achieves the triple $(R, 1 - R, 1)$.

Combining the two parts, the achievable region of the wiretap channel II is obtained. It seems that to attain good tradeoffs, large matrices (large N) are required. In the following sections, the characteristics of the matrices good for the wiretap channel II are investigated. It turns out that there also exist good matrices of small sizes for the channel, but only for certain rates (K/N).

6.4 Equivocation and IDLP

In [9], Forney introduced concepts associated with linear block codes used to analyze their trellis complexity. One of the concepts called Inverse Dimension/Length Profile of an (N, K) linear block code C, demonstrates the dependency of the equivocation on the structure of the matrix $\mathbf{A}$ generating the code. Denoted by $\tilde{k}(C)$, the IDLP of C is defined as a sequence of length $N + 1$ with components

$$\tilde{k}_i(C) = \min_{\tau}\{k[P_\tau(C)] : |\tau| = i\}, 0 \leq i \leq N,$$

where $P_\tau(C)$ is the projection of C onto τ. The projection of a code C onto an index set $\tau \subseteq \{0, 1, \ldots, N\}$ is defined as the code C whose all codewords' components indexed by the members of τ^c are set to 0, and $k[P_\tau(C)]$ is the dimension of the projection.

Denoted by a set consisting of its codewords, a (3,2) code $C = \{000, 011, 101, 110\}$ is used in an example to calculate the sequence $\tilde{k}(C)$. There are four terms in the sequence corresponding to four possible cardinalities of τ. There is one way of selecting τ for $|\tau| = 0$ or 3, and there are three ways of selecting τ for $|\tau| = 1$ or 2. Hence,

$$\begin{aligned}
\tilde{k}_0(C) &= \min\{0\} = 0 \\
\tilde{k}_1(C) &= \min\{1, 1, 1\} = 1 \\
\tilde{k}_2(C) &= \min\{2, 2, 2\} = 2 \\
\tilde{k}_3(C) &= \min\{2\} = 2,
\end{aligned}$$

and $\tilde{k}(C) = \{0, 1, 2, 2\}$. In the case that a code C is generated by a matrix $\mathbf{A}$, $k[P_\tau(C)] = \text{rank}(\mathbf{A}_\tau)$. Then,

$$\tilde{k}_i(C) = \min_\tau \{\text{rank}(\mathbf{A}_\tau) : |\tau| = i\}, 0 \leq i \leq N. \tag{6.5}$$

Equations (6.2) and (6.5) imply that if a matrix $\mathbf{A}$ is used to generate a codebook for the wiretap channel II according to the procedure described in section 6.2, and it is a generator matrix for a code C, then the i^{th} component $\tilde{k}_i(C)$ of the IDLP is the equivocation of the message encoded by the system using the matrix $\mathbf{A}$, given any $N - i$ bits of a transmitted vector being wiretapped. $\tilde{k}(C)$ can be used as a performance index of the code generated from a matrix for the wiretap channel II.

The (3,2) code $C = \{000, 011, 101, 110\}$ is generated from the matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Hence, if the matrix $\mathbf{A}$ is used for generating the codebook for the wiretap channel II, then the $\tilde{k}(C) = \{0, 1, 2, 2\}$ indicates that the equivocations of the message are 0, 1, 2 and 2 if the wiretapper can respectively choose 3, 2, 1, 0 bit(s) from the transmitted vector while the rate is 2/3 with zero probability of error.

6.5 An Upper Bound on Equivocation

The matrix $\mathbf{A}$ generating the codebook for the wiretap channel II determines the performance of the system in that the dimension of $\mathbf{A}$ signifies the rate (K/N), and $\tilde{k}(C)$ indicates the equivocation sequence. Given a fixed rate K/N, a system designer seeks to find a $K \times N$ matrix so that each component of the associated IDLP is as high as possible. An upper bound on IDLP is given by Forney in [9] as

$$\tilde{k}_i(C) \leq UP(\tilde{k}) = \begin{cases} K, & K \leq i \leq N, \\ i, & 0 \leq i \leq K. \end{cases} \tag{6.6}$$

If the upper bound $UP(\tilde{k})$ is written in term of μ $(i = N - \mu)$ and normalized by the message length K, and let $\alpha = \mu/N$, then

$$\begin{aligned} UP(\tilde{k}) &= \begin{cases} K, & 0 \leq \mu \leq N - K \\ N - \mu, & N - K \leq \mu \leq N \end{cases} \\ \frac{UP(\tilde{k})}{K} &= \begin{cases} 1, & 0 \leq \frac{\mu}{N} \leq \frac{N-K}{N} \\ \frac{N-\mu}{K}, & \frac{N-K}{N} \leq \frac{\mu}{N} \leq 1 \end{cases} \\ &= \begin{cases} 1, & 0 \leq \alpha \leq 1 - R \\ \frac{1-\alpha}{R}, & 1 - R \leq \alpha \leq 1 \end{cases} \end{aligned}$$

which coincides with (6.3) – the boundary of the capacity region of the wiretap channel II. Note that $UP(\tilde{k})$ depends on K and N. Hence, at a given rate $R = K/N$, perfect secrecy of the message can be maintained if $\mu \leq N - K$ and there exists a code that achieves the upper bound. However, if $\mu > N - K$, there is no way to maintain perfect secrecy.

If the sender knows exactly the number of bits μ that the wiretapper can choose to observe, he can try to find a matrix such that $\tilde{k}_{N-\mu}(C)$ meets the upper bound. When μ is unknown to the sender, a matrix that allows the IDLP to meets the upper bound at all points is desired.

The sufficient and necessary condition for the IDLP of a code to achieve the upper bound at all points shown in [9] is that the code is MDS, which is defined as an (N, K) code with minimum distance of $N - K + 1$. Unfortunately, there are only 3 types of binary MDS codes:

- $(N, 1)$ code: $\mathbf{A} = [1\ 1\ 1\ \ldots\ 1]$;
- $(N, N-1)$ code: $\mathbf{A} = [\mathbf{u}\ \mathbf{I}]$, where $\mathbf{u}$ is the all-one vector, and $\mathbf{I}$ is the identity matrix;
- (N, N) code: $\mathbf{A} = \mathbf{I}$.

We further note that performing the elementary row operations on the matrix $\mathbf{A}$ does not change its IDLP structure as shown in the example in the following section.

6.6 An Example of Rate 2/3 Binary Codes

Consider the rate 2/3 code $C = \{000, 011, 101, 110\}$ used in section 6.4 generated by the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Since the minimum distance of C is 2 = 3-2+1, C is an MDS code. We also note that $\mathbf{A}$ is also a parity-check matrix of the code $C^{\perp} = \{000, 111\}$. The codebook generated from the matrix $\mathbf{A}$ is

s	transmitted vectors
00	000,111
01	010,101
10	011,100
11	001,110

Table 6.1: Rate 2/3 codebook for the wiretap channel II

In section 6.4, $\tilde{k}(C)$ was calculated to be $\{0, 1, 2, 2\}$. According to equation (6.6), the upper bound on the equivocation for rate 2/3 codes is $\{0, 1, 2, 2\}$. The IDLP of the code C then meets the upper bound, and the code for the wiretap channel II generated from the matrix $\mathbf{A}$ is the most robust rate 2/3 binary code because it provides maximum equivocation for all possible μ.

6.7 A Wiretapper with Side Information

Now that the properties of good binary codes for the wiretap channel II are identified, we investigate the situation, in which the wiretapper has an additional capability of obtaining a

part of the uncoded message as side information. We are interested in how well the coding strategy described in section 6.2 performs in this situation.

Let ρ be an index set such that $\rho \subseteq \{1, 2, \ldots, K\}$ and $|\rho| = a_2$. The wiretapper can select the set ρ to index a_2 components of S^K as side information. Let $S_1^{a_1} = S_{\rho^c}^K$ and $S_2^{a_2} = S_\rho^K$, where $a_1 + a_2 = K$. Under this circumstance, a system designer seeks to maximize the conditional equivocation

$$\Delta_{S_1|S_2,Z} = \min_{\substack{\tau^c:|\tau^c|=\mu \\ \rho:|\rho|=a_2}} H(\mathbf{S}_1^{a_1}|\mathbf{S}_2^{a_2}, \mathbf{Z}^\mu).$$

Employing the same coding strategy with a parity-check matrix $\mathbf{A}$ as before with a more capable wiretapper, the task of the wiretapper in extracting the message becomes simpler. The wiretapper now needs to solve fewer sets of simultaneous equations with fewer unknowns:

$$\begin{bmatrix} \mathbf{A}_{1\tau} \\ \mathbf{A}_{2\tau} \end{bmatrix} \mathbf{x}_\tau^T = \begin{bmatrix} \mathbf{s}_1^T \\ \mathbf{s}_2^T \end{bmatrix} + \begin{bmatrix} \mathbf{A}_{1\tau^c} \\ \mathbf{A}_{2\tau^c} \end{bmatrix} \mathbf{x}_{\tau^c}^T,$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \text{ and } \mathbf{A}_\tau = \begin{bmatrix} \mathbf{A}_{1\tau} \\ \mathbf{A}_{2\tau} \end{bmatrix},$$

and $\text{rank}(\mathbf{A}_1) = a_1$; $\text{rank}(\mathbf{A}_2) = a_2$. The goal of the designer then becomes maximizing

$$\begin{aligned} \Delta_{S_1|S_2,Z} &= \min_{\substack{\tau^c:|\tau^c|=\mu \\ \rho:|\rho|=a_2}} H(\mathbf{S}_1^{a_1}|\mathbf{S}_2^{a_2}, \mathbf{Z}^\mu) \\ &= \min_{\substack{\tau^c:|\tau^c|=\mu \\ \rho:|\rho|=a_2}} \{H(\mathbf{S}_1^{a_1}, \mathbf{S}_2^{a_2}|\mathbf{Z}^\mu) - H(\mathbf{S}_2^{a_2}|\mathbf{Z}^\mu)\} \\ &= \min_{\substack{\tau:|\tau|=N-\mu \\ \rho:|\rho|=a_2}} \{\text{rank}(\mathbf{A}_\tau) - \text{rank}(\mathbf{A}_{2\tau})\}. \end{aligned} \tag{6.7}$$

To study the characteristics of good codes for this wiretap channel II with side information, we make a connection, in the following sections, between the equivocation and the concept of Inverse Relative Dimension/Length Profile (IRDLP) introduced in [15].

6.8 Equivocation and IRDLP

Motivated by the communication situation of the multiuser wiretap channel, the concept of IDLP associated with a linear block code was extended to Inverse Relative Dimension/Length Profile (IRDLP) of a linear block code and its subcodes in [15]. IRDLP of an (N, K) code C^1 and its subcode C^2 is defined as the sequence $\widetilde{K}(C^1, C^2)$ with components $\widetilde{K}_i(C^1, C^2) = \min_\tau\{k[P_\tau(C^1)] - k[P_\tau(C^2)] : |\tau| = i\}, 0 \le i \le N$, where C^2 is an a_2-dimensional linear subspace of C^1.

If the code C^1 is generated by the matrix $\mathbf{A}$, and the code C^2 is generated by the submatrix $\mathbf{A}_2$ defined in the previous section, then the C^2 is a subcode of C^1, and

$$\begin{aligned} \widetilde{K}_i(C^1, C^2) &= \min_{\tau:|\tau|=i} \{k[P_\tau(C^1)] - k[P_\tau(C^2)]\} \\ &= \min_{\tau:|\tau|=i} \{\text{rank}(\mathbf{A}_\tau) - \text{rank}(\mathbf{A}_{2\tau})\} \\ &= \min_{\tau^c:|\tau^c|=N-i} \{H(\mathbf{S}_1^{a_1}, \mathbf{S}_2^{a_2}|\mathbf{Z}^\mu) - H(\mathbf{S}_2^{a_2}|\mathbf{Z}^\mu)\}. \end{aligned}$$

$\widetilde{K}(C^1, C^2)$, the structure of the specific code/subcode pair (C^1, C^2), is the minimization over all possible τ while the equivocation is the minimization over all possible τ and ρ. Consequently, the i^{th} component $\widetilde{K}_i(C^1, C^2)$ of the IRDLP is the equivocation of the information given any $N-i$ bits of a transmitted vector being wiretapped and that $S_2^{a_2}$ corresponding to C^2 is given to the wiretapper as side information. Therefore, IRDLP of a linear block code, C^1, and its subcode, C^2, is the sequence of equivocation of the information when the matrix $\mathbf{A}$ is used to encode S^K, and $\mathbf{A}_2$ corresponds to the side information given to the wiretapper, and μ bits are wiretapped, where μ runs from N down to 0. An upper bound for the IRDLP and characteristics of a code/subcode pair achieving the upper bound are discussed in the next section.

6.9 An Upper Bound on the Performance

In [15], an upper bound on the IRDLP associated with an (N, K) code C^1 and its a_2-dimensional subcode C^2 is proved to be

$$\widetilde{K}(C^1, C^2) \leq UP(\widetilde{K}) = \{0, \ldots, 0, 1, 2, \ldots, K-a_2, \ldots, K-a_2\},$$

where $\max\{i : \widetilde{K}_i(C^1, C^2) = 0\} = a_2$. Equivalently,

$$\widetilde{K}_i(C^1, C^2) \leq UP(\tilde{K}) = \begin{cases} 0, & 0 \leq i \leq a_2, \\ i - a_2, & a_2 \leq i \leq K, \\ K - a_2, & K \leq i \leq N. \end{cases}$$

For $i = N - \mu$,

$$\widetilde{K}_{N-\mu}(C^1, C^2) \leq \begin{cases} 0, & N - a_2 \leq \mu \leq N, \\ N - \mu - a_2, & N - K \leq \mu \leq N - a_2, \\ a_1, & 0 \leq \mu \leq N - K. \end{cases} \tag{6.8}$$

The upper bound indicates that, for rate K/N, and $S_2^{a_2}$ corresponding to C^2 is given to the wiretapper as side information, the system can guarantee that the wiretapper can be ignorant of the information $S_1^{a_1}$ if the number of bits being wiretapped is less than or equal to $N-K$, and there exists a code/subcode pair achieving the upper bound, but nothing can be guaranteed if the number of bits being wiretapped is more than or equal to $N - a_2$.

In the case that the set ρ and the parameter μ are known in advance, the system designer can look for a code/subcode pair (C^1, C^2) such that $\tilde{K}_{N-\mu}(C^1, C^2)$ meets the upper bound $UP(\tilde{K})$ to attain the optimum equivocation. The set ρ and the parameter μ are however not necessarily known to the system designer beforehand. It is therefore desirable to have a code C^1 such that the upper bound is achieved for all possible ρ and μ. The theorem below provides a sufficient condition for a code and its subcode to meet the upper bound on IRDLP.

Theorem 6.5 *If C^1 is an MDS code, then for any C^2 being a subcode of C^1, $\widetilde{K}(C^1, C^2) = UP(\widetilde{K})$.*

Proof: Let C^1 be an (N, K) MDS code under the hypothesis. Let C^2 be an (N, a_2) subcode of C^1 and $\mathbf{A}_2$ be an $a_2 \times N$ generator matrix of C^2. Now, construct a $K \times N$ generator matrix

$\mathbf{A}$ of C^1 by appending $K - a_2$ length-N row vectors to $\mathbf{A}_2$. Because C^1 is an MDS code, $\tilde{k}(C^1) = UP(\tilde{k})$, i.e.,

$$\min_{\tau:|\tau|=i} \text{rank}(\mathbf{A}_\tau) = UP(\tilde{k}) = \begin{cases} i & \text{if } 0 \le i \le K \\ K & \text{if } K < i \le N. \end{cases}$$

Any i columns of the matrix $\mathbf{A}$ are independent for $i \le K$, which implies that rank($\mathbf{A}$) depends only on $|\tau| = i$, and is independent of a specific τ, yielding $\min_{\tau:|\tau|=i} \text{rank}(\mathbf{A}_\tau) = \min\{i, K\}$. Hence,

$$\begin{aligned} \widetilde{K}_i(C^1, C^2) &= \min_{\tau:|\tau|=i} \{\text{rank}(\mathbf{A}_\tau) - \text{rank}(\mathbf{A}_{2\tau})\} \\ &= \begin{cases} i - \max_{\tau:|\tau|=i} \text{rank}(\mathbf{A}_{2\tau}) & \text{if } 0 \le i \le K, \\ K - \max_{\tau:|\tau|=i} \text{rank}(\mathbf{A}_{2\tau}) & \text{if } K \le i \le N \end{cases} \\ &= \begin{cases} i - i & \text{if } 0 \le i \le a_2, \\ i - a_2 & \text{if } a_2 \le i \le K, \\ K - a_2 & \text{if } K \le i \le N. \end{cases} \\ &= UP(\tilde{K}), \end{aligned}$$

and the theorem follows. ■

This theorem suggests that when a generator matrix of an (N, K) MDS code exists and is used to generate a codebook for the wiretap channel II with side information, the codebook is the most robust codebook for rate K/N with zero probability of error since the codebook allows the maximum equivocation for such coding strategy for all possible ρ and μ.

Consider the code $C^1 = \{000, 011, 101, 110\}$ used as a rate 2/3 example in section 6.6. If $|\rho| = a_2 = 1$, then $UP(\tilde{K}) = \{0, 0, 1, 1\}$ for $N = 3, K = 2$ and $a_2 = 1$. Let the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

be the generator matrix of C^1 and be used to generate the codebook for the wiretap channel II with side information. Assume that $\mu = 1$. Two sets of ρ's are possible. If $\rho = \{1\}$, the corresponding submatrix for the side information is $\mathbf{A}_2 = [1\ 0\ 1]$, and the corresponding C^2 is $\{000, 101\}$. As a result,

$$\begin{aligned} \tilde{K}_0(C^1, C^2) &= \min\{0\} = 0 \\ \tilde{K}_1(C^1, C^2) &= \min\{0, 1, 0\} = 0 \\ \tilde{K}_2(C^1, C^2) &= \min\{1, 1, 1\} = 1 \\ \tilde{K}_3(C^1, C^2) &= \min\{1\} = 1, \end{aligned}$$

and $\tilde{K}(C^1, C^2) = \{0, 0, 1, 1\} = UP(\tilde{K})$.

Similarly, if If $\rho = \{2\}$, the corresponding submatrix for the side information is $\mathbf{A}_2 = [0\ 1\ 1]$, and the corresponding C^2 is $\{000, 011\}$. As a result,

$$\begin{aligned} \tilde{K}_0(C^1, C^2) &= \min\{0\} = 0 \\ \tilde{K}_1(C^1, C^2) &= \min\{1, 0, 0\} = 0 \\ \tilde{K}_2(C^1, C^2) &= \min\{1, 1, 1\} = 1 \\ \tilde{K}_3(C^1, C^2) &= \min\{1\} = 1, \end{aligned}$$

and $\tilde{K}(C^1, C^2) = \{0, 0, 1, 1\} = UP(\tilde{K})$. The implication is that the one bit of the message is kept secret from the wiretapper if the other bit of the message and one bit of the codeword are given to the wiretapper.

Although it is shown that using generator matrix of an MDS code C^1 to construct a codebook for the wiretap channel II with side information is sufficient for the IRDLP of a code/subcode pair (C^1, C^2) to meet the upper bound, it is not necessary as shown in the following theorem.

Theorem 6.6 *Let C be an (N, K) binary MDS code generated by a matrix $\mathbf{M}$, and $\mathbf{l}$ be $[1\ 0\ 0\ \ldots\ 0]^T$ of length N. Let C^1 be the code generated by the matrix $\mathbf{A} = [\mathbf{l}\ \mathbf{M}]$ used in the encoding/decoding processes, and C^2 be generated by $\mathbf{A}_2$ being the last $K-1$ rows of $\mathbf{A}$. Then $\widetilde{K}(C^1, C^2) = UP(\widetilde{K})$.*

Proof: By hypothesis of the theorem, $\text{rank}(\mathbf{M}) = K$, and rank of the last $K-1$ rows of $\mathbf{M}$ is $K-1$. $\text{rank}(\mathbf{A}_2) = K-1$ since $\mathbf{A}_2$ is the last $K-1$ rows of $\mathbf{M}$ prepended by a zero vector. Then, $UP(\tilde{K}) = \{0, \ldots, 0, 1, \ldots, 1\}$, where $\max\{i : UP(\widetilde{K})_i = 0\} = K-1$. To compute $\tilde{K}(C^1, C^2)$, first consider the first K terms in the sequence. For $0 \le i \le K-1$, we can select $\tau \subseteq \{2, 3, \ldots, N\}$ such that $|\tau| = i$ and $k[P_\tau(C^1)] - k[P_\tau(C^2)] = i - i = 0$ since C is MDS; hence, $\widetilde{K}_i(C^1, C^2) = 0$ for $0 \le i \le K-1$. To compute the $i^{th} = K^{th}$ term in the sequence, we consider the following three cases.

- case 1: $1 \notin \tau$,
  - $k[P_\tau(C^1)] - k[P_\tau(C^2)] = K - k[P_\tau(C^2)] \ge 1$ since C is MDS;
- case 2: $1 \in \tau$ and $k[P_{\tau_0}(C^1)] - k[P_{\tau_0}(C^2)] = 0$ where $\tau_0 = \tau \setminus \{1\}$,
  - $k[P_{\tau_0}(C^2)] = k[P_{\tau_0}(C^1)] = K-1$ since C is MDS;
  - $k[P_\tau(C^2)] = k[P_{\tau_0}(C^2)] = K-1$ by construction;
  - $k[P_\tau(C^1)] = k[P_\tau(C^2)] + 1 = K$ by construction;
  - hence, $k[P_\tau(C^1)] - k[P_\tau(C^2)] = 1$;
- case 3: $1 \in \tau$ and $k[P_{\tau_0}(C^1)] - k[P_{\tau_0}(C^2)] = 1$ where $\tau_0 = \tau \setminus \{1\}$,
  - $k[P_\tau(C^1)] - k[P_\tau(C^2)] = k[P_\tau(C^1)] - k[P_{\tau_0}(C^2)]$ by construction;
  - $k[P_\tau(C^1)] - k[P_{\tau_0}(C^2)] \ge k[P_{\tau_0}(C^1)] - k[P_{\tau_0}(C^2)] = 1$;
  - $k[P_\tau(C^1)] - k[P_\tau(C^2)] \le 1$;
  - hence, $k[P_\tau(C^1)] - k[P_\tau(C^2)] = 1$.

Hence, $\widetilde{K}_K(C^1, C^2) = 1$. For $K < i \le N$, $\widetilde{K}(C^1, C^2)$ is a non-decreasing sequence (from [15]), and $\widetilde{K}_i(C^1, C^2) \le UP(\widetilde{K}) \Rightarrow \widetilde{K}_i(C^1, C^2) = 1$, and the theorem follows. ■

As an example, let's construct a codebook for the wiretap channel II with side information based on Theorem 6.6. Let C be the MDS code $\{000, 011, 101, 110\}$ and $\rho = \{2\}$. Then the matrix $\mathbf{A}$ is constructed to be

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

s	transmitted vectors
00	0000,0111,1100,1011
01	0010,0101,1110,1001
10	0011,0100,1000,1111
11	0001,0110,1010,1101

Table 6.2: Rate 1/2 codebook for the wiretap channel II with side information.

and the corresponding $\mathbf{A}_2$ is [0 0 1 1]. Then, $C^1 = \{0000, 1101, 0011, 1110\}$, and $C^2 = \{0000, 0011\}$, and the codebook constructed based on the matrix $\mathbf{A}$ is

The terms in the IRDLP sequence are

$$\begin{aligned}
\tilde{K}_0(C^1, C^2) &= \min\{0\} = 0 \\
\tilde{K}_1(C^1, C^2) &= \min\{1, 1, 0, 0\} = 0 \\
\tilde{K}_2(C^1, C^2) &= \min\{1, 1, 1, 1, 1, 1\} = 1 \\
\tilde{K}_3(C^1, C^2) &= \min\{1, 1, 1, 1\} = 1 \\
\tilde{K}_4(C^1, C^2) &= \min\{1\} = 1,
\end{aligned}$$

and $\tilde{K}(C^1, C^2) = \{0, 0, 1, 1, 1\} = UP(\tilde{K})$. However, for $\rho = \{1\}$, $\mathbf{A}_2 = [1\ 1\ 0\ 1]$, and $C^2 = \{0000, 1101\}$. The terms in the IRDLP sequence are

$$\begin{aligned}
\tilde{K}_0(C^1, C^2) &= \min\{0\} = 0 \\
\tilde{K}_1(C^1, C^2) &= \min\{0, 0, 1, 0\} = 0 \\
\tilde{K}_2(C^1, C^2) &= \min\{0, 1, 1, 1, 1, 1\} = 0 \\
\tilde{K}_3(C^1, C^2) &= \min\{1, 1, 1, 1\} = 1 \\
\tilde{K}_4(C^1, C^2) &= \min\{1\} = 1,
\end{aligned}$$

and $\tilde{K}(C^1, C^2) = \{0, 0, 0, 1, 1\} \leq \{0, 0, 1, 1, 1\} = UP(\tilde{K})$.

If the codebook in Table 6.2 is employed, the rate of transmission is 1/2, and the equivocation sequence meets the upper bound $UP(\tilde{K})$ for all possible μ when $\rho = \{2\}$. However, for $\rho = \{1\}$, the equivocation does not meet the upper bound for $\mu = 2$. This codebook makes the first bit of the message more vulnerable. In contrast to using this codebook, the codebook in Table 6.1 allows the equivocation to meet the upper bound for all possible ρ and μ.

From the examples above, the coset coding method based on linear block codes allows perfect secrecy for the communication through the wiretap channel II with/without side information. Nevertheless, to achieve the perfect secrecy, there are certain strong restrictions on the rate of transmission and the difference in the secret sensitivity of the message bits to the parameter ρ. If arbitrarily small deviation from the perfect secrecy is allowed, it is shown in the next section that there exist good linear block codes for the coset coding method at all rates smaller than one.

6.10 An Achievable Region

In this section, an achievable region for the wiretap channel II with side information is proved based on the same strategy as that used in [16]. A quadruple (R, R_1, α, δ) is said to be achievable if, for all $\epsilon < 0$, there exists an encoder/decoder pair with parameters $N \geq N_0$ such that

$$\begin{aligned} K &\geq (R-\epsilon)N, \\ a_1 &\geq (R_1-\epsilon)N, \\ \mu &\geq (\alpha-\epsilon)N, \\ \Delta_{S_1|S_2,Z} &\geq (\delta-\epsilon)a_1 \\ P_e &\leq \epsilon, \end{aligned}$$

where

$$\begin{aligned} \Delta_{S_1|S_2,Z} &= \min_{\substack{\tau^c:|\tau^c|=\mu \\ \rho:|\rho|=a_2}} H(S_1^{a_1}|S_2^{a_2}, Z^\mu) \\ P_e &= \frac{1}{K}\sum_{k=1}^{K} \Pr\{S_k \neq \hat{S}_k\}, \end{aligned}$$

K is length of the message, a_2 is length of the side information given to the wiretapper, $a_1 = K - a_2$, and μ is the number of wiretapped bits.

To prove the achievable region of the wiretap channel II with side information, we make a direct use of the properties of binary random matrices as in the proof of the capacity region of the wiretap channel II (Lemmas 6.16 and 6.17) and extend Lemma 6.15 to Lemma 6.18 in order to accommodate the side information.

Lemma 6.18 *Suppose that a parity-check matrix $\mathbf{A}_{K\times N}$ is used with the coset coding method for the wiretap channel II with side information, and there are two wiretappers who have parameters $(\tau_1^c, \Delta_{S_1|S_2,Z_1})$ and $(\tau_2^c, \Delta_{S_1|S_2,Z_2})$, where τ_1^c and τ_2^c are the index sets used to identify the components of the transmitted vectors observed by the wiretappers. If $|\tau_2^c| = \mu_2 \geq |\tau_1^c| = \mu_1$, then*

$$\Delta_{S_1|S_2,Z_2} \geq \Delta_{S_1|S_2,Z_1} - (\mu_2 - \mu_1).$$

Proof: From equation (6.7), the equivocation is related to the matrix $\mathbf{A}$ by

$$\Delta_{S_1|S_2,Z} = \min_{\substack{\tau:|\tau|=N-\mu \\ \rho:|\rho|=a_2}} \{\mathrm{rank}(\mathbf{A}_\tau) - \mathrm{rank}(\mathbf{A}_{\rho\tau})\}.$$

Let τ_1, τ_2 and ρ_1 ρ_2 respectively be the index sets to the codeword components and message components so that $|\tau_1^c| = \mu_1$, $|\tau_2^c| = \mu_2$, $|\rho_1| = |\rho_2| = a_2$ and

$$\begin{aligned} \Delta_{S_1|S_2,Z_1} &= \mathrm{rank}(\mathbf{A}_{\tau_1}) - \mathrm{rank}(\mathbf{A}_{\rho_1\tau_1}) \\ \Delta_{S_1|S_2,Z_2} &= \mathrm{rank}(\mathbf{A}_{\tau_2}) - \mathrm{rank}(\mathbf{A}_{\rho_2\tau_2}). \end{aligned}$$

Let's consider the relation of the equivocations in two cases: when $\tau_2 \subseteq \tau_1$ and when $\tau_2 \not\subseteq \tau_1$.

case 1: $\tau_2 \subseteq \tau_1$
Consider the matrices $\mathbf{A}_{\tau_2}$ and $\mathbf{A}_{\rho_2\tau_2}$, with the fact that $\tau_2 \subseteq \tau_1$, we have the relations:

$$\begin{aligned} \operatorname{rank}(\mathbf{A}_{\tau_1}) - (\mu_2 - \mu_1) \leq \quad & \operatorname{rank}(\mathbf{A}_{\tau_2}) \quad \leq \operatorname{rank}(\mathbf{A}_{\tau_1}) && (6.9) \\ \operatorname{rank}(\mathbf{A}_{\rho_2\tau_1}) - (\mu_2 - \mu_1) \leq \quad & \operatorname{rank}(\mathbf{A}_{\rho_2\tau_2}) \quad \leq \operatorname{rank}(\mathbf{A}_{\rho_2\tau_1}). && (6.10) \end{aligned}$$

Hence,

$$\begin{aligned} \Delta_{S_1|S_2,Z_2} = \operatorname{rank}(\mathbf{A}_{\tau_2}) - \operatorname{rank}(\mathbf{A}_{\rho_2\tau_2}) &\geq \operatorname{rank}(\mathbf{A}_{\tau_1}) - (\mu_2 - \mu_1) - \operatorname{rank}(\mathbf{A}_{\rho_2\tau_1}) \\ &\geq \operatorname{rank}(\mathbf{A}_{\tau_1}) - (\mu_2 - \mu_1) - \operatorname{rank}(\mathbf{A}_{\rho_1\tau_1}) \\ &= \Delta_{S_1|S_2,Z_1} - (\mu_2 - \mu_1). \end{aligned}$$

case 2: $\tau_2 \not\subseteq \tau_1$
Let $\tau^* \subseteq \{1, 2, \ldots, N\}$ be an index set to columns of $\mathbf{A}$ such that $\tau_2 \cap \tau^* = \emptyset$ and $|\tau^*| = \mu_2 - \mu_1$. Since appending columns of $\mathbf{A}$ indexed by τ^* to the matrix $\mathbf{A}_{\tau_2}$ increases the equivocation at most $\mu_2 - \mu_1$,

$$\begin{aligned} \Delta_{S_1|S_2,Z_2} + (\mu_2 - \mu_1) &\geq \operatorname{rank}(\mathbf{A}_{\tau_2\cup\tau^*}) - \operatorname{rank}(\mathbf{A}_{\rho_2(\tau_2\cup\tau^*)}) \\ &\geq \operatorname{rank}(\mathbf{A}_{\tau_1}) - \operatorname{rank}(\mathbf{A}_{\rho_1\tau_1}) = \Delta_{S_1|S_2,Z_1} \\ \Delta_{S_1|S_2,Z_2} &\geq \Delta_{S_1|S_2,Z_1} - (\mu_2 - \mu_1). \end{aligned}$$

Combining the two cases results in the theorem. ■

We are now ready to prove the achievable region for the wiretap channel II with side information based on the coset coding method with linear block codes in the following Theorem.

Theorem 6.7 *If the coset coding method is used with linear block codes for the wiretap channel II with side information , (R, R_1, α, δ) is achievable if $R \geq 0, 0 \leq R_1 \leq R, \alpha \leq 1$ and*

$$0 \leq \delta \leq \begin{cases} 1, & 0 \leq \alpha \leq 1 - R, \\ \frac{1-\alpha-(R-R_1)}{R_1}, & 1 - R \leq \alpha \leq 1 - (R - R_1), \\ 0, & 1 - (R - R_1) \leq \alpha \leq 1. \end{cases}$$

Proof: Given $R \geq 0$, $0 \leq R_1 \leq R$, we first show that $(R, R_1, 1 - R, 1)$ is achievable for all $\epsilon > 0$. The rest follows from Lemma 6.18.

Let $\mathbf{A}$ be a $K \times N$ matrix to be used with the coset coding method for the wiretap channel II with side information. The rate of transmission then is K/N and $R_1 = (K - a_2)/N$. For a given ϵ, let $\mu = N - K - \epsilon N$ so that $\alpha = \mu/N = 1 - R - \epsilon$. We will show that there exists at least a matrix that can be used as the matrix $\mathbf{A}$ so that $\delta \geq 1 - \epsilon$.

Define an indicator function of the matrix $\mathbf{A}$ as

$$\Psi(\mathbf{A}) = \begin{cases} 1 & \text{if } \min_{\substack{\tau:|\tau|=N-\mu \\ \rho:|\rho|=a_2}} \{\operatorname{rank}(\mathbf{A}_\tau) - \operatorname{rank}(\mathbf{A}_{\rho\tau})\} < a_1 - L \text{ or } \operatorname{rank}(\mathbf{A}) < K, \\ 0 & \text{otherwise,} \end{cases}$$

for $0 \leq L \leq a_1$, where $a_1 = K - a_2$ and $\mathbf{A}_{\rho\tau}$ is the submatrix of $\mathbf{A}$ consisting of rows indexed by members of ρ and columns indexed by members of τ. We must then show that there exists a matrix $\mathbf{A}$ with $\Psi(\mathbf{A}) = 0$.

$\Psi(\mathbf{A})$ is upper bounded by

$$\Psi(\mathbf{A}) \leq \sum_{\substack{\tau:|\tau|=N-\mu \\ \rho:|\rho|=a_2}} \Phi(\mathbf{A},\tau,\rho) + \Phi_0(\mathbf{A}),$$

where

$$\Phi_0(\mathbf{A}) = \begin{cases} 1 & \text{if } \mathrm{rank}(\mathbf{A}) < K, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\Phi(\mathbf{A},\tau,\rho) = \begin{cases} 1 & \text{if } \mathrm{rank}(\mathbf{A}_\tau) - \mathrm{rank}(\mathbf{A}_{\rho\tau}) < a_1 - L, \\ 0 & \text{otherwise.} \end{cases}$$

Now if the matrix $\mathbf{A}_{K\times N}$ is selected at random from all possible 2^{KN} matrices, we have, for given τ such that $|\tau| = N-\mu$, ρ such that $|\rho| = a_2$, $a_1 = K - |\rho|$ and $0 \leq L \leq a_1$,

$$\begin{aligned} \Pr\{\mathrm{rank}(\mathbf{A}_\tau) - \mathrm{rank}(\mathbf{A}_{\rho\tau}) < a_1 - L\} &= \Pr\{\mathrm{rank}(\mathbf{A}_\tau) < a_1 - L + \mathrm{rank}(\mathbf{A}_{\rho\tau})\} \\ &\leq \Pr\{\mathrm{rank}(\mathbf{A}_\tau) < a_1 - L + a_2\} \\ &= \Pr\{\mathrm{rank}(\mathbf{A}_\tau) < K - L\} \\ &\leq 2^{N-\mu} 2^{-(L+1)(N-\mu-K)}. \end{aligned}$$

The last inequality follows from applying Lemma 6.16 to the $K \times (N-\mu)$ matrix $\mathbf{A}_\tau$. For $\Phi(\mathbf{A},\tau,\rho)$ is an indicator function, $E[\Phi(\mathbf{A},\tau,\rho)] = \Pr\{\mathrm{rank}(\mathbf{A}_\tau) - \mathrm{rank}(\mathbf{A}_{\rho\tau}) < a_1 - L\} \leq 2^{N-\mu} 2^{-(L+1)(N-\mu-K)}$.

Furthermore, from Lemma 6.17,

$$E[\Phi_0(\mathbf{A})] = \Pr\{\mathrm{rank}(\mathbf{A}) < K\} \leq \frac{K 2^{K-1-N}}{1-2^{K-1-N}} \leq \frac{K 2^{K-N}}{1-2^{K-N}}.$$

Since there are no more than 2^N possible τ's and 2^N possible ρ's,

$$E[\Psi(\mathbf{A})] \leq 2^{2N} 2^{-(L+1)(N-\mu-K)+(N-\mu)} + \frac{K 2^{K-N}}{1-2^{K-N}}$$

For $\mu = N - K - \epsilon N$ and $K = RN$,

$$E[\Psi(\mathbf{A})] \leq 2^{-L\epsilon N + 3N} + \frac{RN 2^{-(1-R)N}}{1-2^{-(1-R)N}}.$$

For sufficiently large N_1 and $L \geq 4/\epsilon$, $E[\Psi(\mathbf{A})] < 1$. Since $\Psi(\cdot)$ is an integer-valued function, there must exist a $K \times N$ matrix $\mathbf{A}^*$ such that $\Psi(\mathbf{A}^*) = 0$ so that

$$\begin{aligned} \mathrm{rank}(\mathbf{A}^*) &= K, \text{and} \\ a_1\delta &= \Delta_{S_1|S_2,Z} \\ &= \min_{\substack{\tau:|\tau|=N-\mu \\ \rho:|\rho|=a_2}} \{\mathrm{rank}(\mathbf{A}^*_\tau) - \mathrm{rank}(\mathbf{A}^*_{2\tau})\} \\ &\geq a_1 - L. \end{aligned}$$

For any $\epsilon > 0$ and N_0, pick $L \geq 4/\epsilon$, $N \geq \max\{N_0, N_1, L/R_1\epsilon\}$. Then $\delta \geq 1 - L/R_1 N \geq 1 - \epsilon$. ■

Under the same assumption of using the coset coding method with linear block codes, the converse to Theorem 6.7 is also true. That is if (R, R_1, α, δ) is achievable, $R \geq 0$, $0 \leq R_1 \leq R$, $0 \leq \alpha \leq 1$, and

$$\delta \leq \begin{cases} 1, & 0 \leq \alpha \leq 1-R, \\ \frac{1-\alpha-(R-R_1)}{R_1}, & 1-R \leq \alpha \leq 1-(R-R_1), \\ 0, & 1-(R-R_1) \leq \alpha \leq 1. \end{cases}$$

since

$$\begin{aligned} \Delta_{S_1|S_2,Z} &= \min_{\substack{\tau^c:|\tau^c|=\mu \\ \rho:|\rho|=a_2}} H(S_1^{a_1}, S_2^{a_2}|Z^\mu) - H(S_2^{a_2}|Z^\mu) \\ &= \min_{\substack{\tau:|\tau|=N-\mu \\ \rho:|\rho|=a_2}} [\mathrm{rank}(\mathbf{A}_\tau) - \mathrm{rank}(\mathbf{A}_{2\tau})] \\ &\leq \begin{cases} a_1, & 0 \leq \mu \leq N-K, \\ N-\mu-a_2, & N-K \leq \mu \leq N-a_2, \\ 0, & N-a_2 \leq \mu \leq N. \end{cases} \end{aligned}$$

due to equation (6.7) and the upper bound (6.8).

6.11 Remarks

We have seen that the generator matrices of MDS codes can be used in conjunction with the coset coding method to construct the most robust codes for the wiretap channel II as suggested by the linkage between the performance of the coding strategy and the associated Inverse Dimension/Length Profile. Furthermore, applying the analysis of the extended concept of Inverse Relative Dimension/Length Profile to the wiretap channel II with side information suggests that the generator matrices of MDS codes can be used to construct the most robust codes for this model, too. However, there are other types of codes that perform as robust but require the knowledge of the side information at the encoder as show in Theorem 6.6.

An achievable region for the wiretap channel II with side information is derived based on the coset coding method with linear block codes. However, we believe that it is possible to enlarge the region if the coding strategy is not restricted to the coset coding method based on linear block codes.

Chapter 7

Concluding Remarks

7.1 Summary of the Work

Motivated by the covert communication scenario, the code-partitioning technique is closely examined in an information theoretic framework. In particular, it can be used to accomplish different objectives for different communication situations. For the broadcast channel, it is used for simultaneous transmission of different messages to different recipients. For the wiretap channel, it is used to keep the secret message from the adversary. For the dirty-paper channel, it is used to mitigate the effect of the interference in the channel known to the sender in advance.

The new model of the Gaussian wiretap channel with side information is introduced as a theoretical base for the covert communication application. The model is analyzed based on the leakage function, which can be used to classify the channel. The camouflage mode of operation is more power-effective, but it is only suitable for some channels. The high-power mode uses the input power to confuse the adversary and to send the message at the same time and is less power-effective. Based on the proposed coding strategy, an achievable region is proved.

The model is adapted for information embedding in an image by considering the image as an interference in the model known in advance to the sender. We also extend a quantization-based watermarking scheme developed from the dirty-paper channel model. Two extensions are proposed to enhance the secrecy of the message embedded in the image. The direct approach enhances the secrecy of the message at the cost of the degradation of the image while the indirect approach enhances the secrecy of the message at the cost of much smaller degradation.

The question of the use of side information by the adversary is investigated under the wiretap channel II with side information model based on the optimal coding strategy proposed for the wiretap channel II. The performance of the codes are linked to the structures called IDLP and IRDLP, suggesting the characteristics of good codes for the wiretap channel II and the wiretap channel II with side information. An achievable region for the wiretap channel II with side information is then proved based on the coding strategy.

7.2 Possible Directions for Further Work

The Gaussian wiretap channel with side information model is a promising candidate for the information hiding application as illustrated in Chapter 5 even though the implementation is not fully optimized for the secrecy of the message. The capacity region of the Gaussian wiretap channel with side information remains unknown. A theoretical challenge is to enlarge the achievable region or finding the capacity region.

On the application side, we have illustrated that a simple implementation of the idea is possible by the direct and indirect approaches of extension. However, the fundamental differences between the two approaches have not been analyzed. Further analysis is necessary to improve and fine tune the performance of the purposed implementation. In addition, the question of secrecy when the key sequence is used repeatedly has not been coped with.

Finally, the analysis of the coset coding method based on linear block codes provides an insight into the use of matrices in constructing robust codes for the wiretap channel II with and without side information. However, the capacity region of the wiretap channel II with side information still remains unknown. Allowing other types of codes may be able to enlarge the achievable region.

Bibliography

[1] Patrick P. Bergmans. Random coding theorem for broadcast channels with degraded components. *IEEE Transactions on Information Theory*, IT-19(2):197–207, March 1973.

[2] Patrick P. Bergmans. A simple converse for broadcast channels with additive white gaussian noise. *IEEE Transactions on Information Theory*, pages 279–280, March 1974.

[3] Max H. M. Costa. Writing on dirty paper. *IEEE Transactions on Information Theory*, IT-29(3):439–441, May 1983.

[4] Thomas M. Cover. Broadcast channels. *IEEE Transactions on Information Theory*, IT-19(1):2–14, January 1972.

[5] Thomas M. Cover. Comments on broadcast channels. *IEEE Transactions on Information Theory*, IT-44(6):2524–2530, October 1998.

[6] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., N. Y., 1991.

[7] Wilbur B. Davenport, Jr. *Probability and Random Processes*. McGraw-Hill Book Company, Inc., 1970.

[8] J. J. Eggers, J. K. Su, and B. Girod. A blind watermarking scheme based on structured codebooks. *Secure Images and Image Authentication, IEE Colloquium*, pages 4/1–4/6, April 2000.

[9] G. David Forney, Jr. Dimension/length profiles and trellis complexity of linear block codes. *IEEE Transactions on Information Theory*, 40(6):439–441, November 1994.

[10] R. G. Gallager. *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., N. Y., 1965.

[11] R. G. Gallager. Capacity and coding for degraded broadcast channels. *Probl. Inform. Transm.*, pages 185–193, July-Sept. 1974.

[12] S. I. Gel'fand and M. S. Pinsker. Coding for channel with random parameters. *Problems of Control and Information Theory*, 9(1):19–31, 1980.

[13] Steven R. Lay. *Analysis with an Introduction to Proof, Second Edition*. Prentice-Hall, Inc., New Jersey, 1990.

[14] S. K. Leung-Yan-Cheong and Martin E. Hellman. The gaussian wire-tap channel. *IEEE Transactions on Information Theory*, IT-24(4):451–456, July 1978.

[15] Yuan Luo, Chaichana Mitrpant, and A. J. Han Vinck. The multi-user wire-tap channel of type ii using coset coding method. 2003.

[16] L. H. Ozarow and A. D. Wyner. Wire-tap channel ii. *AT&T Bells Laboratories Technical Journal*, 63(10):2135–2157, December 1984.

[17] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, July 1948.

[18] M. R. Spiegel. *Schaum's Mathematical Handbook*. McGraw-Hill Book Company, Inc., 1968.

[19] John M. Wozencraft and Irwin Mark Jacobs. *Principles of Communication Engineering*. John Wiley & Sons, Inc., 1965.

[20] A. D. Wyner. The wire-tap channel. *The Bell System Technical Journal*, 54(8):1355–1387, October 1975.

Appendix A

Background Theorems

The Weak Law of Large Number

Theorem A.1 *Let X^N be a sequence of independent identically distributed random variables with mean m and variance σ_X^2, for any $\epsilon > 0$,*

$$Pr\left\{ \left| \frac{1}{N} \sum_{i=1}^{N} X_i - m \right| < \epsilon \right\} \quad > \quad 1 - \frac{\sigma_X^2}{N\epsilon^2}. \tag{A.1}$$

Proof: See [19]. ■

The Union Bound

Theorem A.2 *Let $A_i, i \in \{1, \ldots, N\}$ be events of interest of an experiment. Then*

$$Pr\left\{ \bigcup_{i=1}^{N} A_i \right\} \leq \sum_{i=1}^{N} Pr\{A_i\}.$$

Proof: See [7] and [19]. ■

Fano's Inequality

Theorem A.3 *For discrete random variables $X, Y, \hat{X}$ such that $X \rightarrow Y \rightarrow \hat{X}$ forms a Markov chain. Define the probability of error $P_e = Pr\{\hat{X} \neq X\}$. Then $h(P_e) = P_e \log(|\mathcal{X}| - 1) \geq H(X|Y)$.*

Proof: See [6]. ■

Appendix B

Abbreviations and Notations

Abbreviations

AEP	Asymptotic equipartition properties
AWGN	Additive white Gaussian noise
BSC	Binary symmetric channel
DMC	Discrete memoryless channel
GWC	Gaussian wiretap channel
GWCSI	Gaussian wiretap channel with side information
i.i.d.	Independent identically distributed
IDLP	Inverse Dimension/Length Profile
IRDLP	Inverse Relative Dimension/Length Profile
WLLN	Weak law of large numbers
WT2CSI	Wiretap channel II with side information

Notations

$\alpha, \delta, \epsilon, \omega, \xi$	constants
(a, b)	open interval between a and b
$[a, b]$	closed interval between a and b
$\mathbf{A}$	matrix
$\mathbf{A}_{\rho\tau}$	submatrix of $\mathbf{A}$ whose columns are indexed by members of τ, whose rows are indexed by members of ρ
$\|\mathbf{A}\|$	determinant of $\mathbf{A}$
C	channel capacity
C_{AWGN}	capacity of AWGN channel
$C^1, C^2, C^{aux}, C^{sat}$	codes
Δ	equivocation
$E[X]$	expectation of X
$h(p)$	binary entropy function with cross-over probability p
$H(X)$	(differential) entropy of X
$H(X, Y)$	joint (differential) entropy of X and Y
$I(X; Y)$	mutual information between X and Y
$k[C]$	dimension of a code C
$\tilde{k}(C)$	IDLP of a code C
$\widetilde{K}(C^1, C^2)$	IRDLP of a code/subcode pair (C^1, C^2)
$\mathbf{K}$	covariance matrix
p	cross-over probability
$p(x)$	probability mass (density) function
$p(x, y)$	joint probability mass (density) function
$p(x\|y)$	conditional probability mass (density) function
$P_\tau(C)$	projection of a code C on to an index set τ
$\mathcal{Q}(\cdot)$	quantization operator
ϱ	correlation coefficient
σ_X^2	variance of X
τ, ρ	index sets
τ^c	complement of a set
$T_X^N(\epsilon)$	set of typical sequences of length N with respect to a probability density function $p(x)$ and a constant ϵ
$T_{X,Y}^N(\epsilon)$	set of pairs of typical sequences of length N with respect to a joint probability density function $p(x, y)$ and a constant ϵ
$\mathcal{U}$	quantization set
$UP(\tilde{k})$	upper bound on IDLP
$UP(\widetilde{K})$	upper bound on IRDLP
$\mathbf{x}$	vector
x^N	vector of length N
$\mathbf{x}^T$	transpose of vector $\mathbf{x}$
$X, X_c, \tilde{X}, \hat{X}, \chi$	random variables
$X \sim \mathcal{N}(0, \sigma_X^2)$	X has normal distribution with mean zero and variance σ_X^2
$\mathcal{X}$	set of all possible alphabets of X
$\emptyset$	empty set

List of Figures

List of Tables